Essays on the Economics of Betting Markets

Alasdair Brown

Department of Financial and Management Studies

School of Oriental and African Studies, University of London

A thesis submitted to fulfill the requirements for the degree of

Doctor of Philosophy

February 2012

Abstract

This thesis comprises three sections. The first utilises novel data from a recently developed betting exchange to examine informational issues in market microstructure. Among other things, the betting exchange allows us to observe (uncontrolled) private information and learn about the behaviour of informed traders in a simple asset market. The second section studies U.K. horse-racing for signs of conflict between horse owners (principals) and trainers (agents). Trainers often prepare their own horses for races in addition to having outsiders' horses in their care. We find evidence of agent shirking, as observed in other arenas, but also evidence of an informational rent - linked to betting market manipulation - extracted by the agent. The third section takes a microscopic look at market efficiency and the limits to arbitrage on a betting exchange. The arbitrage trade on which we focus is not subject to the most prominent limit - noise-trader risk - which gives us a cleaner examination of the remaining limits to arbitrage.

Contents

Acknowledgements

I would like to thank my supervisors, Fabrizio Adriani and Pasquale Scaramozzino, for their extensive advice on this thesis. I learnt a great deal under their guidance but remain responsible for all errors.

The papers in this thesis have benefited from the review process at *Applied Economics, Economic Inquiry, The European Journal of Finance,* and *The International Journal of Forecasting.* I would like to thank the editors of these journals, and a number of anonymous referees. This thesis has also been improved by audiences at my old home of SOAS, my new home at the University of East Anglia, and attendees at seminars at the Royal Economic Society Easter School (Birmingham, 2010), the French Finance Association meeting (St. Malo, 2010), Warwick Business School and Cardiff Business School.

Finally, I would like to thank my mother Rajeswary, my father Ian, and my brother Andrew for the financial, technical, and crucially, emotional support they have given me during this time.

Chapter 1

Introduction

The majority of economics papers on betting markets have focused on market efficiency (see Vaughan Williams (2005) for a survey). As betting markets are short-lived, yield easily-quantifiable final payoffs for the assets traded, and have a degree of repetition, they have provided clean tests of efficiency. In this thesis, we utilise betting markets - some new and some more established - to study three other fields in economics.

This is the first piece of work, as far as we are aware, to use betting exchanges to explicitly analyse informational issues in market microstructure. In Chapter 2 we examine the effect of market segmentation on price discovery. Specifically, do informed bettors choose new markets which allow them to exploit the greater precision in their information, or traditional markets which offer greater liquidity? In Chapter 3, we consider the nature of informed trading on a betting exchange; in particular, do the informed derive their advantage from private information or from a superior ability to process public information?

In Chapter 4, we use betting market data to consider a fundamental question in economics; namely, what effect does the separation of ownership and control have on the payoff of the owner? We examine agency conflict in horse-racing, where many trainers (agents) divide their time between the preparation of their own horses and those of outsiders (principals). While the trainer receives all of the win purse if their horse is successful, they receive only a fraction (typically 10%) for training an outsider's horse to victory. We examine whether the agent shirks in this environment and, further, whether the temptation of the accompanying betting market affects the behaviour of the agent.

Finally, in Chapter 5, we return to market efficiency to consider a descendant of this field - the limits to arbitrage. Recent work has focused on how these limits explain market inefficiencies (see Gromb and Vayanos (2010) for a survey). We take an arbitrage trade that is not subject to noise-trader risk and examine if arbitrage opportunities still arise. If they do, which remaining limits to arbitrage are responsible?

References

- Gromb, D., Vayanos, D., (2010). Limits of Arbitrage: The State of the Theory. Paul Woolley Centre Working Paper.
- Vaughan Williams, L., (2005), *Information Efficiency in Financial and Betting Markets*, Cambridge University Press.

Chapter 2

The Importance of Liquidity: Price Discovery in Segmented Markets on a Betting Exchange

Abstract

Online bettors can now bet on sporting events via the traditional win bets, or with a variety of new exotic bets, on which they wager on outcomes other than the winner of the event. We propose that this presents informed bettors with a conundrum: trade in the exotic market, where their precise private information is of most value, or trade in the more liquid win market, where they can obtain better prices. Using in-play betting data from the latter stages of The Wimbledon Tennis Championships in 2009, we find that the price discovery contribution of exotic markets is negligible, suggesting that the informed sacrifice a proportion of their private information in return for the liquidity, and superior price execution, of the main win market. This result suggests that, in general, creating more markets does not necessarily improve price discovery.

2.1 Introduction

Recent years have seen an explosion in the number of exotic betting markets, where bettors wager on outcomes other than the winner of a sporting event. Punters can, among other things, bet on the number of goals in a football match, the score in sets of a tennis match, and the score from each ball in a cricket match. This development has taken place at the same time as the growth of betting exchanges, which have increased competition and reduced transaction costs, and the parallel rise of in-play betting, where bettors can place bets on the outcome of a sporting event as it unfolds.

The most extreme exotic markets have recently attracted bad publicity. In August 2010, the Pakistan cricket team was caught up in an alleged 'spot-fixing' scandal, broken by the U.K. tabloid newspaper, *The News of the World*. In this episode, a number of players were accused of agreeing to bowl no-balls - where the bowler oversteps their mark and forfeits a run to the opposing side - at pre-specified times, in return for cash from a fake betting syndicate. This scandal followed a similar episode of alleged spot-fixing in English county cricket in April of the same year (*The Telegraph*, May 31st 2010).[1]

Nevertheless, some exotic bets may allow informed bettors to exploit the greater precision of their legally obtained information. As this information is revealed through betting, this may aid the price discovery process. To support this proposition, *The New York Times* and *The Guardian* (June 24th 2009) document an instance of bettors trad-

[1]An earlier instance of the dubious use of exotic betting markets has been found in football. In his autobiography, *Taking Le Tiss*, the Southampton player Matt Le Tissier admits betting on the time of the first throw-in in a match in April 1995. The bet in this case was a spread bet, where a bettor is liable for the degree to which they are wrong with their prediction. Having predicted an early throw-in, Le Tissier planned to kick the ball straight out from the kick-off. The following passage describes how his teammate Neil Shipperley unwittingly sabotaged his bet. 'As it was live on TV, I didn't want to make it too obvious or end up looking like a prat for miscuing the ball, so I tried to hit it just over his head. But, with so much riding on it, I was a bit nervous and didn't give it enough welly. The problem was that Shipps knew nothing about the bet and managed to reach it and head it back into play. I have never run so much in my life..as I charged round the pitch desperately trying to kick the ball out of play.' The ball eventually went out after 70 seconds, meaning no money was won or lost.

ing on private information on a tennis match in the first round of the Men's Wimbledon Championships. The Austrian Jurgen Melzer was drawn to play the American Wayne Odesnik. Prior to the match, information leaked on a minor injury to Odesnik. As a result, a number of insiders bet large quantities, not on Melzer to win, but on Melzer to win 3-0 (the eventual score), thereby leveraging their precise information.

The market microstructure literature does, however, suggest that there is a trade-off.[2] If informed bettors were to disproportionately congregate in the exotic market, then the liquidity in this market, and their ability to execute their bets at advantageous prices, would diminish. As the proportion of informed in a trading population increases, those providing liquidity will widen the bid-ask spread (e.g. Copeland and Galai (1983) and Glosten and Milgrom (1985)), and limit the volume available to trade (e.g. DuPont 2000). In this chapter, we seek to ascertain which effect dominates. Do the informed trade in the exotic market, where their precise information is of most value, or does the lower liquidity in exotic markets force them back into the main win market, where they can hide amongst the uninformed?

In this study we examine in-play betting, on a betting exchange, on a series of tennis matches in the 2009 Wimbledon tournament. In tennis betting markets, a bettor can replicate a bet on a player to win with a series of bets in the set market. By constructing a replicating portfolio, we can examine whether price discovery takes place in the win market or in the exotic set market. By using a similar methodology to Garbade and Silber (1983) and Hasbrouck (1995), we are able to calculate the information share of each type of market. We find that the win market provides most, if not all, of the price discovery in tennis markets. In other words, the informed sacrifice a proportion of their information in return for the liquidity of the win market and the superior price execution it promises. This result suggests that, in general, creating more markets does not necessarily improve price discovery.

[2]The market microstructure literature examines the effect of the trading mechanism - and economic frictions such as asymmetric information and liquidity-control - on prices in financial markets. Prominent surveys include O'Hara (1995), Madhavan (2000), Biais, Glosten and Spatt (2005) and Hasbrouck (2007). For a focus on limit order books, see Parlour and Seppi (2008).

Betting markets have long provided an interesting setting for studies of market efficiency.[3] In contrast to many financial assets, betting assets are short-lived, allowing bettors to hold profitable assets until maturity. This circumvents many of the limits to arbitrage present in financial markets.[4] The structure of the betting exchange in our study is almost identical to that found on most financial exchanges. The exchange operates as a continuous double auction, with long or short positions possible, and the option of market orders (executed at prices in the book) or limit orders (which sit in the book until an offsetting order arrives).

The rest of the chapter is structured as follows. Section 2.2 describes the related empirical literature in financial and betting markets. Section 2.3 outlines the structure of the betting exchange and the data, while Section 2.4 contains the empirical model and results. Section 2.5 concludes.

2.2 Related Literature

The ideas covered in this chapter relate to three strands of the financial economics literature. The first strand looks at price discovery in segmented markets. It is possible to trade the same asset in a number of geographical regions and to gain an exposure to the same collection of cash flows through a variety of different asset structures (e.g. stocks, options, futures etc.). A literature has therefore emerged to locate where price discovery takes place, and by implication, where the informed trade.

Stocks can be traded in 'dominant' markets and in 'satellite' markets. Hasbrouck (1995) looks at the relative contribution of the New York Stock Exchange and regional exchanges - such as those found in Boston, Philadelphia and Cincinnati - to price discovery. Using a methodology developed along the lines of Garbade and Silber (1983), the author calculates the information share of each exchange. Among the thirty Dow

[3] For a survey of the betting market efficiency literature, see Vaughan Williams (2005) and Hausch and Ziemba (2008).

[4] The presence of noise-traders, who may drive prices further from fundamentals, constitutes a risk to arbitrage in most financial markets. For a survey of this and other limits to arbitrage, see Gromb and Vayanos (2010).

stocks, the median contribution of the NYSE to price discovery is found to be 92.7%.

Other studies have looked at the relative contributions of stock and option markets to price discovery. Starting with Black (1975), it has been argued that informed traders will favour the option market, where their precise information can be leveraged with small initial capital outlays. Chakravarty *et al* (2004) modify Hasbrouck's (1995) methodology to capture the arbitrage relationships between stock and option markets. Using a data set of stock and option prices for 60 firms over 5 years, they find that the option market contributes around 17% of price discovery. The authors also found that the contribution of option markets was highest in out-of-the-money options, where they were making highly leveraged, and precise, predictions about future stock price movements. Other authors, including Easley *at al* (1998), have also tied the volume found in option markets to the presence of informed traders.

The second strand of the literature examines the relationship between liquidity and market efficiency. In an effort to explain the post-earnings announcement drift - which shows that asset prices do not completely and instantaneously incorporate all information contained in announcements - Chordia *et al* (2006) examine the drift for liquid and illiquid stocks. The authors find that illiquid stocks display a greater drift than liquid stocks (1.79% compared to 0.24%) and that most of the returns on offer are absorbed by the transaction costs of illiquidity. Chordia *et al* (2008) look at the more general short-run predictability of returns against liquidity. Again, illiquid stocks display the greatest predictability, suggesting that liquid markets are quicker to incorporate information and attain efficiency. Tetlock (2008) examines the relationship between liquidity and market efficiency on a betting exchange, TradeSports, similar to that used in our study. In contrast to the earlier results in financial markets, the author finds that liquid markets are no bettor predictors of event outcomes, i.e. no more efficient, than illiquid markets.[5]

[5] Other studies of betting exchange efficiency include Smith *et al* (2006) and Tetlock (2004), who examine the favourite-longshot bias, and Zitzewitz (2006), who looks at the information contained in short-term binary financial options. Spann and Skiera (2009) compare the predictive capacity of a prediction market with that of professional tipsters and bookmakers.

There is a third related strand of the literature, focusing on the efficiency of exotic betting markets. Comparing the returns to win bets and the returns of a replicating portfolio composed of exotic bets, in horse-racing, the following studies focus on the pari-mutuel market, where the payout is inversely proportional to the amount wagered on an outcome. Ali (1979) examines the efficiency of win and exotic markets from 1079 observations from U.S. racetracks in 1975, and finds that win and exotic markets efficiently price equivalent bets. On the other hand, Asch and Quandt (1987) - using data from 705 races at Meadowland racetrack in 1984 - find inefficiencies, albeit inefficiencies which do not yield a profit to the bettor once the track take is acknowledged. For further examples of these studies, see Hausch *et al* (2008).

With respect to the efficiency of tennis betting markets, a recent paper by Easton and Uylangco (2009) uses betting exchange data to compare the point-by-point pricing of a player to win, with that suggested in a theoretical point-by-point model. They find that the (main win) betting market is relatively efficient, and in some cases, correctly anticipates the significance of the outcome of a service game. Other features, such as the tendency of players to lose a number of points after having their serve broken, are less well captured. The efficiency of tennis betting markets, with respect to expected returns, is also considered in Forrest and McHale (2007).

2.3 The Exchange

In our study we used betting price data from Betfair, the most prominent betting exchange in the U.K.. Betfair provides a public limit order book for the exchange of bets on a variety of sporting events. The book operates as a continuous double auction, with prices quoted in the form of back odds (where the bettor places a stake and receives the stake multiplied by the odds if the outcome occurs) and lay odds (where the stake is accepted and the bettor is liable for the stake multiplied by the odds if the outcome occurs). As with other exchanges, bettors can place market orders, at prices currently in the book, or limit orders, which reside in the book until an offsetting order arrives. As there is no designated market-maker, limit orders represent the sole provision of liquidity on the exchange.

For tennis matches, there are bets traded on the winner of the match, and also the score in sets. Both of these markets operate during play. This allows us to continuously compare the implied probability of a player winning, as defined by the win market, with the implied probability defined by the exotic set market. Suppose that there were back and lay odds of B_W and L_W respectively for a player to win. For each 1 unit staked (or backed) on this outcome, B_W would be returned in addition to the stake if the player wins. For each unit laid, L_W would be lost if the player wins. Taking the midpoint of the spread as the asset's value, these odds imply the probability of the player winning is

$$W = \frac{\frac{1}{B_W+1} + \frac{1}{L_W+1}}{2} \tag{2.1}$$

At the same time, suppose the set market quotes back (lay) odds of B_0 (L_0), B_1 (L_1) and B_2 (L_2) on the same player to win 3-0, 3-1 and 3-2 respectively. This implies the probability of the player winning is

$$S = \frac{\frac{1}{B_0+1} + \frac{1}{L_0+1}}{2} + \frac{\frac{1}{B_1+1} + \frac{1}{L_1+1}}{2} + \frac{\frac{1}{B_2+1} + \frac{1}{L_2+1}}{2} \tag{2.2}$$

For our analysis we examined 7 matches from the latter stages of the Men's Wimbledon Tennis Championships in 2009, as these matches provide the most liquid markets. This data was obtained from Fracsoft, who market Betfair historical pricing data. This data includes the in-play quoted back and lay odds, and associated volumes, for each player in a match to win, and also the score in sets. Using this data we calculated the implied probabilities, W and S, as defined above.

Table 2.1 describes the summary statistics on liquidity for 14 players in those 7 matches. Liquidity is on average lower in the set market than the win market. This is reflected by wider back-lay spreads and lower volumes. The microstructure literature predicts that liquidity will decrease as the proportion of informed in a trading population increases. Our statistics suggest that liquidity providers expect that the informed disproportionately congregate in the exotic set market, to fully exploit the greater precision of their information. Those providing liquidity therefore increase the cost of trading in this market to limit their losses. In the next section we analyse whether this cost is sufficient to deter the informed.

Match	Player	IP	Win Volume	Set Volume	Win Spread	Set Spread
Roddick Federer	1	0.12	1930.18	136.19	0.0083	0.0349
	2	0.88	61828.55	349.78	0.0071	0.0300
Murray Roddick	1	0.72	18924.43	214.03	0.0072	0.0305
	2	0.28	6024.76	164.41	0.0075	0.0268
Federer Hass	1	0.92	387540.39	403.25	0.0088	0.0231
	2	0.08	975.21	44.83	0.0029	0.0174
Karlovic Federer	1	0.11	600.58	90.12	0.0028	0.0242
	2	0.9	333568.71	357.85	0.0090	0.0532
Roddick Hewitt	1	0.72	36478.41	289.12	0.0072	0.1071
	2	0.28	1953.24	112.66	0.0149	0.0988
Djokovic Hass	1	0.69	5222.11	234.49	0.0124	0.0986
	2	0.31	12932.82	281.29	0.0082	0.0641
Murray Ferrero	1	0.89	133966.01	411.14	0.0077	0.0428
	2	0.11	522.13	44.96	0.0050	0.0385

Table 2.1. Liquidity: Summary Statistics. This table describes the starting implied probability, and the average (back-lay) spread and volume in both the win and set markets, for 14 players to win in 7 different matches, from the quarter-final stages onwards of the Men's Wimbledon Tennis Championships in 2009. Volume in the set market is the summation of volume for all three bets. Liquidity is lower, in the form of wider spreads and lower volumes, in the set market for all 14 players.

2.4 Empirical Analysis

The model of price discovery described in this chapter is based on the model of Garbade and Silber (1983), who looked at the relative price discovery of spot and futures markets in a number of commodities. Underpinning this model is the idea that any divergence in prices of identical assets in segmented markets will be rapidly arbitraged.

To confirm this in our setting, we looked at the duration of arbitrage opportunities between the win and exotic set market on Betfair. Chapter 5 provides a detailed discussion of arbitrage strategies and the remaining limits to arbitrage on Betfair. A bettor can arbitrage price differences by taking a long (short) position in the win (set) market and taking an offsetting short (long) position in the set (win) market. In Table

2.2 we describe summary statistics of the duration of arbitrage opportunities for these two strategies, for the 14 players of our study.

Match	Player	Strategy	mean	s.d.	med.	max.	IQR	Freq.	N
Roddick Federer	1	1	7.51	11.97	4.00	67	6.00	6.45	5032
		2	4.67	2.08	4.00	7	2.00	0.28	5032
	2	1	3.28	2.74	2.00	9	3.00	2.46	2395
		2	9.50	11.54	4.50	36	4.75	3.97	2395
Murray Roddick	1	1	12.25	12.69	7.00	31	9.75	2.20	2259
		2	8.93	10.52	3.50	37	13.25	5.53	2259
	2	1	6.15	8.21	3.00	26	4.00	1.88	4333
		2	14.29	20.09	8.00	73	10.75	4.62	4333
Federer Hass	1	1	5.10	3.86	4.00	13	3.25	1.50	7370
		2	7.31	5.55	6.50	22	5.50	1.60	7370
	2	1	7.00	6.98	6.50	14	11.50	1.11	2805
		2	NA	NA	NA	NA	NA	0.00	2805
Karlovic Federer	1	1	7.00	NA	7.00	7	0.00	1.35	1352
		2	NA	NA	NA	NA	NA	0.00	1352
	2	1	15.25	10.53	16.00	27	10.75	2.03	6186
		2	3.67	1.15	3.00	5	1.00	0.18	6186
Roddick Hewitt	1	1	7.95	6.01	6.00	21	6.50	6.42	5334
		2	2.50	0.71	2.50	3	0.50	0.09	5334
	2	1	1.00	NA	1.00	1	0.00	26.09	1646
		2	NA	NA	NA	NA	NA	0.00	1646
Djokovic Hass	1	1	12.00	NA	12.00	12	0.00	1.61	2744
		2	NA	NA	NA	NA	NA	0.00	2744
	2	1	4.33	3.21	3.00	8	3.00	0.24	7879
		2	11.19	15.64	5.50	73	10.75	3.69	7879
Murray Ferrero	1	1	8.19	12.23	3.50	63	8.25	3.90	6140
		2	8.09	5.47	7.00	19	6.00	1.49	6140
	2	1	NA	NA	NA	NA	NA	0.00	2726
		2	27.06	38.02	6.00	144	48.00	16.87	2726

Table 2.2. Arbitrage Duration: Summary Statistics. This table describes summary statistics on the duration of arbitrage strategies for the 14 players, in 7 matches, of the Men's Wimbledon Tennis Championships in 2009. Strategy 1 (2) consists of betting on (against) a player to win in the win market, and betting against (on) the same player to win in the exotic set market. Durations are calculated whilst all set outcomes are possible, and a 2% commission on net winnings is assumed. For a detailed discussion of these strategies, and arbitrage on Betfair, see Chapter 6.

We observe that arbitrage opportunities are frequent during play, occuring in all 7 matches studied, and can persist, lasting for 144 seconds in the case of Ferrero strategy 2. However, most opportunities are quickly removed, with the median duration of arbitrage opportunities below 10 seconds for all but two of the samples. Figure 2.1 provides further evidence of the close pricing relationship between the win market and the exotic set market. Taking data from the Final itself, it appears that arbitrage pressures ensure that any pricing deviations between the win and set markets are short-lived. For this series, and for all others subsequently considered, we also conducted non-parametric Phillips-Perron tests to confirm that the difference in win and set prices was stationary. The presence of a unit root was rejected in all cases so this allows for the use of a Garbade and Silber-type model.[6]

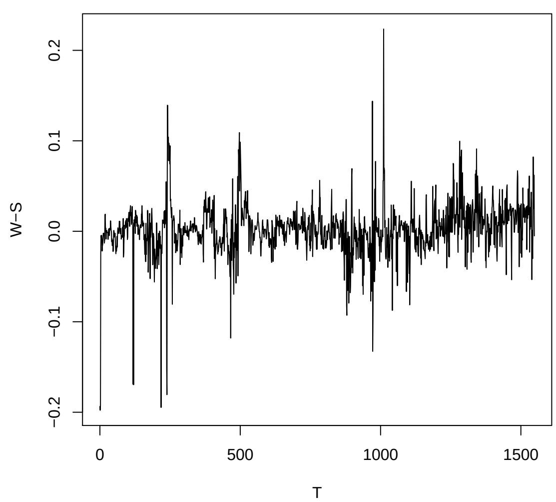

Figure 2.1: The difference in the probability of a Roger Federer win in the 2009 Wimbledon Final, as implied by the win market, Equation (2.1), and the set market, Equation (2.2). Prices sampled at 10 second intervals during the match.

For our analysis we estimated the following two regressions:

$$W_t - W_{t-1} = \beta_0 + \beta_1(S_{t-1} - W_{t-1}) + \epsilon_t \tag{2.3}$$

$$S_t - S_{t-1} = \alpha_0 + \alpha_1(W_{t-1} - S_{t-1}) + u_t \tag{2.4}$$

[6]The Garbade and Silber (1983) model differs from Hasbrouck (1995) in that only 2 price series, rather than n, are sampled, and that only one lag is used. As our setting utilises only two prices, and because arbitrage opportunities disappear after one lag, we use the simpler Garbade and Silber framework.

where W_t is the implied probability of a player winning in the win market, as defined in Equation (2.1), S_t is the implied probability of the same player winning in the exotic set market, as defined in Equation (2.2), and ϵ_t and u_t are error terms. The idea is to examine how the two markets respond to a divergence in implied probabilities. For example, if a divergence $(W_{t-1} - S_{t-1})$ causes a subsequent shift in the set market implied probability $(S_t - S_{t-1})$, we can conclude that the win market has hosted the price discovery, and the set market responds to arbitrage pressures. On the other hand, if the set market hosts all the price discovery, we would expect that the divergence $(W_{t-1} - S_{t-1})$ is followed by a change in the win market implied probability $(W_t - W_{t-1})$. Using the estimated results of Equations (2.3) and (2.4), the contribution of the win market to price discovery is defined as

$$\frac{\alpha_1}{\beta_1 + \alpha_1} \tag{2.5}$$

and the contribution of the set market to price discovery is defined as

$$\frac{\beta_1}{\beta_1 + \alpha_1} \tag{2.6}$$

To illustrate the model, consider an instance where the exotic set market implies a probability of 30% for a player winning, whilst the win market implies a 20% probability. If the win market provides all of the price discovery, we would expect that, in one period's time, the set market would also imply a 20% win probability. In this case, the informed are all trading in the win market, and the set market follows the win market due to arbitrage pressures. On the other hand, suppose we estimate that the set market provides 40% of the price discovery. Taking the same example, we would expect both markets to reflect an implied probability of 24% in the next time period. If this were the case, we would expect that some informed bettors were using the set market to trade.

In Table 2.3 we outline the estimation of Equations (2.3) and (2.4) for our 14 players. We estimated the two equations with sampling at intervals of 10 seconds to reflect the brevity of arbitrage opportunities outlined in Table 2.2. We found that the contribution of the exotic set market to price discovery was not significant for 12 of the 14 players.

Of the other two, the set market was found to contribute 1.4%, and 58.5%. The outlier, in this case, is the 58.5% found in the market for Federer to beat Hass in the semi-finals of the tournament. Observing the measures of liquidity described in Table 2.1, we can see that the set market on Federer to beat Hass displayed greater liquidity - in the form of higher volume and narrower spreads - than most of the set markets. With a Federer win a near inevitability (with a pre-match probability of 92%), this suggests that those providing liquidity in the set market did not see this market to be purely the preserve of the informed. With the cost of trading here therefore lowered, a subset of the informed appear to have chosen the exotic market for this particular match.

Match	Player	β_0	β_1	$Adj.R^2$	α_0	α_1	$Adj.R^2$	N	Exotic Cont.
Roddick Federer	1	-0.0001	0.0299	0.0055	0.0016	0.3209	0.1600	1544	0.085
		(0.000338)	(0.019252)		(0.000585)**	(0.068396)***			
	2	-0.0001	-0.0179	0.0006	-0.0014	0.5291	0.2400	1548	-0.035
		(0.000332)	(0.023239)		(0.000696)*	(0.055127)***			
Murray Roddick	1	-0.0003	0.0065	-0.0007	-0.0030	0.3233	0.1626	1125	0.020
		(0.000436)	(0.016510)		(0.000860)***	(0.042034)***			
	2	0.0005	-0.0123	-0.0002	0.0042	0.3762	0.1949	1125	-0.034
		(0.000439)	(0.019080)		(0.000936)***	(0.044229)***			
Federer Hass	1	-0.0020	0.0951	0.0367	0.0006	0.0676	0.0532	736	0.585*
		(0.001316)	(0.044349)*		(0.000459)	(0.035496).			
	2	-0.0001	-0.0043	0.0008	0.0015	0.2328	0.1173	736	-0.019
		(8.11E-05)	(0.003584)		(0.000598)*	(0.071488)**			
Karlovic Federer	1	-0.000212	0.0020	0.0009	0.0031	0.1450	0.0684	618	0.014*
		(8.65E-05)*	(0.000958)*		(0.000883)***	(0.037602)***			
	2	-0.0014	0.0182	0.0053	0.0090	0.1994	0.0897	618	0.084
		(-0.001214)	(0.014413)		(0.003148)**	(0.050296)***			
Roddick Hewitt	1	0.0002	-0.0010	-0.0007	-0.0022	0.1943	0.0940	1385	-0.005
		(0.000446)	(0.004205)		(0.002280)	(0.047630)***			
	2	-0.0001	0.0021	-0.0006	-0.0012	0.1625	0.0792	1385	0.013
		(0.000432)	(0.003886)		(0.001536)	(0.053059)**			
Djokovic Hass	1	-0.0005	0.0031	-0.0007	-0.0106	0.2619	0.1314	987	0.012
		(0.000588)	(0.004557)		(0.002437)***	(0.038705)***			
	2	0.0008	0.0058	0.0001	-0.0074	0.4005	0.2022	987	0.014
		(0.000520)	(0.004453)		(0.003126)*	(0.063062)***			
Murray Ferrero	1	-0.0002	-0.0110	0.0039	0.0030	0.1150	0.0621	614	-0.106
		(0.000447)	(0.010368)		(0.001675).	(0.046058)*			
	2	-0.0002	0.0063	0.0016	-0.0015	0.2541	0.1233	614	0.024
		(0.000189)	(0.004851)		(0.001158)	(0.057625)***			

Table 2.3. Price Discovery: Results. This table describes the results of the estimation of Equations (2.3) and (2.4). These equations examine the contribution to price discovery of the win market and the exotic set market. For this analysis, pricing samples were taken at intervals of 10 seconds for the full duration of each match. Newey-West heteroskedasticity and autocorrelation -consistent standard errors are in parentheses and .,*,**,*** indicates significance at the 10%, 5%, 1% and 0.1% levels respectively. The price discovery contribution of the exotic set market, as defined in Equation (2.6), is presented in the final column.

As mentioned in the introduction, informed traders can use exotic betting markets to leverage their precise information. However, the decline in liquidity that follows informed traders - as counterparties seek to limit their losses with wider spreads and lower volumes - appears to dominate the decision of where to trade. If exotic markets do not contribute to price discovery, the question remains as to their practical value.

One concern may be the omission of autoregressive terms from Equations (2.3) and (2.4). It is possible that these may be confounding factors, and that price changes are temporally dependent. This may cause us to apportion too much weight to price discovery observed in the win market. We therefore ran the supplementary regressions, of Equations (2.7) and (2.8), that follow this paragraph. As illustrated in Table 2.4, the autoregressive terms are predominantly statistically insignificant, and the win market is still the prime driver of price discovery.

$$W_t - W_{t-1} = \beta_0 + \beta_1(S_{t-1} - W_{t-1}) + \beta_2(W_{t-1} - W_{t-2}) + \epsilon_t \tag{2.7}$$

$$S_t - S_{t-1} = \alpha_0 + \alpha_1(W_{t-1} - S_{t-1}) + \alpha_2(S_{t-1} - S_{t-2}) + u_t \tag{2.8}$$

Further, up until this point we have analysed the contribution of exotic markets to price discovery for the whole duration of each match. This includes times when there are three possible scores by which a player could win, but also times when a player can only win by two scores (when they have lost a set), and by only one score (when they have lost two sets). To ensure the robustness of our results we re-estimated Equations (2.3) and (2.4), breaking down the contribution of the exotic market into when there were three, two, and one set score(s) possible. Perhaps the degree of precision required to pick between three scores is too high for all but the most informed, but the precision required to distinguish between two possible scores is within reach of a larger proportion of bettors. If this allows the informed to blend in with others in exotic markets, this may increase the proportion of price discovery that originates here.

Match	Player	β_0	β_1	β_2	$Adj.R^2$	α_0	α_1	α_2	$Adj.R^2$	N	Exotic Cont.
Roddick Federer	1	-0.0001	0.0273	-0.0345	0.0060	0.0016	0.3183	-0.0093	0.1595	1543	0.079
		(0.000348)	(0.019291)	(0.034301)		(0.000608)**	(0.063482)***	(0.058400)			
	2	-0.0001	-0.0237	-0.0839	0.0070	-0.0015	0.5153	-0.0712	0.2518	1543	-0.048
		(0.000355)	(0.023916)	(0.030533)**		(0.000698)*	(0.053537)***	(0.066421)			
Murray Roddick	1	-0.0002	0.0186	0.1291	0.0143	-0.0028	0.2848	-0.1620	0.1858	1124	0.061
		(0.000399)	(0.016310)	(0.039570)**		(0.000833)***	(0.036609)***	(0.057278)**			
	2	0.0004	-0.0066	0.0481	0.0011	0.0042	0.3773	0.0038	0.1943	1124	-0.018
		(0.000434)	(0.018859)	(0.044270)		(0.000955)***	(0.048768)***	(0.054787)			
Federer Hass	1	-0.0021	0.1043	0.0630	0.0390	0.0005	0.0650	-0.1358	0.0703	735	0.616
		(0.001350)	(0.048382)*	(0.026008)*		(0.000474)	(0.036173).	(0.082843)			*
	2	-0.0001	-0.0050	-0.0825	0.0063	0.0014	0.2318	-0.0045	0.1161	735	-0.022
		(8.57E-05)	(0.003888)	(0.067359)		(0.000628)*	(0.072506)**	(0.075127)			
Karlovic Federer	1	-0.0002	0.0021	0.0718	0.0044	0.0028	0.1257	-0.1235	0.0810	617	0.016
		(7.54E-05)**	(0.000993)*	(0.086117)		(0.000853)**	(0.033842)***	(0.066111).			*
	2	-0.0014	0.0183	0.0059	0.0038	0.0084	0.1842	-0.0678	0.0923	617	0.090
		(0.001217)	(0.014452)	(0.006373)		(0.003116)**	(0.050523)***	(0.057083)			
Roddick Hewitt	1	0.0002	-0.0010	-0.0070	-0.0014	-0.0018	0.1657	-0.1519	0.1144	1384	-0.006
		(0.000450)	(0.004240)	(0.023777)		(0.002257)	(0.044171)***	(0.043050)***			
	2	-0.0001	0.0021	-0.0037	-0.0013	-0.0011	0.1365	-0.1733	0.1062	1384	0.015
		(0.000433)	(0.003937)	(0.032703)		(0.001491)	(0.045183)**	(0.064718)**			
Djokovic Hass	1	-0.0006	0.0025	-0.0341	-0.0005	-0.0094	0.2274	-0.1391	0.1476	986	0.011
		(0.000607)	(0.004702)	(0.027385)		(0.002326)***	(0.033615)***	(0.077984).			
	2	0.0008	0.0054	-0.0192	-0.0006	-0.0072	0.3930	-0.0196	0.2017	986	0.013
		(0.000528)	(0.004522)	(0.034035)		(0.003180)*	(0.067125)***	(0.037724)			
Murray Ferrero	1	-0.0002	-0.0118	-0.0226	0.0027	0.0025	0.1054	-0.1358	0.0762	613	-0.126
		(0.000434)	(0.011535)	(0.017377)		(0.001829)	(0.043778)**	(0.093601)			
	2	-0.0002	0.0063	0.0035	0.0000	-0.0015	0.2521	-0.0076	0.1219	613	0.025
		(0.000188)	(0.004849)	(0.036154)		(0.001148)	(0.060715)***	(0.072604)			

Table 2.4. Price Discovery: Results (including autoregressive terms). This table describes the results of the estimation of Equations (2.7) and (2.8). These equations examine the contribution to price discovery of the win market and the exotic set market. For this analysis, pricing samples were taken at intervals of 10 seconds for the full duration of each match. Newey-West heteroskedasticity and autocorrelation -consistent standard errors are in parentheses and .,*,**,*** indicates significance at the 10%, 5%, 1% and 0.1% levels respectively. The price discovery contribution of the exotic set market, as defined in Equation (2.6), is presented in the final column.

Table 2.5 outlines the results of this estimation. The contribution of the exotic set market is highest when there are three outcomes, with two player markets displaying a significant contribution. When a player loses a set and there are only two possible outcomes, the set market does not provide a significant contribution to price discovery for any of the matches. This result suggests that providing less extreme exotic markets, which more closely resemble the main win market, would not lead to a greater hosting of price discovery.

		Exotic Cont.	Exotic Cont.	Exotic Cont.	Exotic Cont.
Match	Player	All	3 outcomes	2 outcomes	1 outcome
Roddick Federer	1	0.0852	0.0699	0.0582	0.1147
	2	-0.035	0.0194	-0.1188	0.1026
. Murray Roddick	1	0.020	0.0433	0.0431	0.0062
	2	-0.034	0.1252	-0.0714	
Federer Hass	1	0.585*	0.5845*		
	2	-0.019	-0.0220	-0.0107	0.0645
Karlovic Federer	1	0.014*	0.0745.	0.0089	0.0026
	2	0.084	0.0835		
Roddick Hewitt	1	-0.005	-0.0054	0.0059	0.0067
	2	0.013	0.0059	0.0254	0.0243
Djokovic Hass	1	0.012	-0.0622	0.0361	0.0274
	2	0.014	0.0112	0.0542	
Murray Ferrero	1	-0.106	-0.1057		
	2	0.024	0.0215	0.0500	0.0474

Table 2.5. Price Discovery: Exotic Market Contribution (Number of Outcomes). This table describes the contribution of the exotic set market to price discovery (as defined in Equation (2.6)), for the whole match, and when broken down into times when there are three possible set outcomes, two possible set outcomes (when a set has been lost), and one possible set outcome (when two sets have been lost). ., *, **, and *** indicate significance at the 10%, 5%, 1% and 0.1% level respectively.

In our analysis thus far, we have also assumed that all divergences between the win and set market are reconciled within 10 seconds. Perhaps arbitrageable divergences are, but smaller divergences take longer to disappear. Bettors must post collateral to bet in separate win and set markets, and therefore some opportunities may be left on the table until a larger opportunity arises. A second problem arises in estimating Equations (2.3) and (2.4) in high frequency data. If trade is infrequent, it is possible that our estimates of β_1 and α_1 are downwardly biased. The downward bias may be most pronounced in the set market if trade is particularly infrequent there. To check the robustness of our results to these possibilities, we re-estimated Equations (2.3) and (2.4), sampling at intervals of 30 seconds and 60 seconds. The results of these estimations are shown in Tables 2.6 and 2.7.

We found our results were robust to these possibilities. When sampling takes place at intervals of 30 seconds, the exotic set market still only provides a significant contribution to price discovery in two instances. These contributions total 1.7% and 69.8%. Again, it is the market on Federer to beat Hass that constitutes the outlier. For sampling at 60 second intervals, there is no significant contribution from the set market. Perhaps more importantly, the sum of the coefficients β_1 and α_1 is increasing in the sampling interval, therefore making it less likely that infrequent trade is driving our results.

In summary, we find that the contribution of the exotic set market to price discovery is negligible. If informed bettors are expected to congregate in a particular market - where their precise information can be leveraged - then liquidity decreases. This decline in liquidity appears to deter the informed and force them back into the main win market, where they can operate under the camouflage of less informed bettors. As our results are found in a relatively benign form of exotic market, where only three exotic bets are needed to replicate a win bet, we would not expect to see improvements in price discovery for the extreme spot-fixing exotic markets mentioned in the news recently.

It should be noted that the framework used in our analysis does not allow for the inventory considerations of liquidity providers. We neglect such considerations because we do not have data on the orders taken by individual liquidity providers, and looking at liquidity provision at a market-wide level may lead to inaccurate conclusions. This is because many of the bookmakers use betting exchanges to offset exposure taken in betting shops and on their own websites. Therefore, what may appear to be a liquidity provider taking a large position may actually be a hedge. Nevertheless, should the necessary data become available, further work could incorporate inventory considerations to address the potential limitations of our analysis.

Match	Player	β_0	β_1	$Adj.R^2$	α_0	α_1	$Adj.R^2$	N	Exotic Cont.
Roddick Federer	1	0.0000	0.0119	-0.0012	0.0031	0.6086	0.3859	516	0.019
		(0.000981)	(0.015182)		(0.001617).	(0.119501)***			
	2	-0.0001	0.0025	-0.0019	-0.0014	0.7360	0.2782	516	0.003
		(0.000988)	(0.044368)		(0.001468)	(0.069753)***			
Murray Roddick	1	-0.0007	0.0649	0.0032	-0.0045	0.5125	0.2095	375	0.112
		(0.001286)	(0.047386)		(0.001996)*	(0.077116)***			
	2	0.0005	0.0667	0.0037	0.0069	0.5966	0.2404	375	0.101
		(0.001162)	(0.035120).		(0.002050)***	(0.092480)***			
Federer Hass	1	-0.0060	0.2994	0.1231	0.0008	0.1295	0.1098	245	0.698
		(0.004540)	(0.118091)*		(0.001004)	(0.076264).			*
	2	-0.0003	-0.0074	-0.0015	0.0037	0.5735	0.2824	245	-0.013
		(0.000216)	(0.008303)		(0.001642)*	(0.133423)***			
Karlovic Federer	1	-0.0006	0.0063	0.0026	0.0079	0.3717	0.1734	206	0.017
		(0.000305).	(0.003508)*		(0.002481)**	(0.100404)***			*
	2	-0.0041	0.0539	0.0146	0.0173	0.3982	0.1687	206	0.119
		(0.003419)	(0.048823)		(0.007723)*	(0.116088)***			
Roddick Hewitt	1	0.0004	-0.0081	-0.0011	-0.0030	0.3449	0.1605	461	-0.024
		(0.001311)	(0.011466)		(0.004868)	(0.081499)***			
	2	-0.0005	-0.0006	-0.0022	-0.0027	0.2881	0.1319	461	-0.002
		(0.001297)	(0.009771)		(0.003549)	(0.092202)**			
Djokovic Hass	1	-0.0025	-0.0145	-0.0003	-0.0210	0.5415	0.2680	329	-0.028
		(0.001703)	(0.015071)		(0.005412)***	(0.093687)***			
	2	0.0025	0.0212	0.0027	-0.0153	0.7859	0.3719	329	0.026
		(0.001639)	(0.018111)		(0.006076)*	(0.091371)***			
Murray Ferrero	1	-0.0018	-0.0022	-0.0049	0.0083	0.2468	0.1618	204	-0.009
		(0.002526)	(0.003565)		(0.003849)*	(0.092705)**			
	2	-0.0005	0.0154	0.0010	-0.0035	0.5160	0.2513	204	0.029
		(0.000632)	(0.015157)		(0.002578)	(0.125124)***			

Table 2.6. Price Discovery: Results (30s intervals). This table describes the results of the estimation of Equations (2.3) and (2.4). These equations examine the contribution to price discovery of the win market and the exotic set market. For this analysis, pricing samples were taken at intervals of 30 seconds for the full duration of each match. Newey-West heteroskedasticity and autocorrelation -consistent standard errors are in parentheses and .,*,**,*** indicates significance at the 10%, 5%, 1% and 0.1% levels respectively. The price discovery contribution of the exotic set market, as defined in Equation (2.6), is presented in the final column.

Match	Player	β_0	β_1	$Adj.R^2$	α_0	α_1	$Adj.R^2$	N	Exotic Cont.
Roddick Federer	1	0.0001	0.0097	-0.0037	0.0030	0.9244	0.5896	258	0.010
		(0.001939)	(0.036046)		(0.002120)	(0.077408)***			
	2	-0.0001	-0.0647	-0.0001	0.0000	0.9065	0.2928	258	-0.077
		(0.001953)	(0.068190)		(0.002633)	(0.081680)***			
Murray Roddick	1	-0.0022	0.0448	-0.0035	-0.0066	0.5296	0.1713	187	0.078
		(0.002397)	(0.098979)		(0.003419).	(0.148516)***			
	2	0.0031	-0.0550	-0.0031	0.0103	0.6542	0.1828	187	-0.092
		(0.002832)	(0.104482)		(0.004378)*	(0.156003)***			
Federer Hass	1	-0.0069	0.2407	0.0861	0.0022	0.0906	0.1031	122	0.726
		(0.006233)	(0.208085)		(0.001527)	(0.070609)			
	2	-0.000508	-0.01624	0.00367	0.006947	0.834016	0.409713	122	-0.020
		(0.000395)	(0.016701)		(0.003216)*	(0.107058)***			
Karlovic Federer	1	-0.0012	0.0107	-0.0012	0.0113	0.5618	0.2541	103	0.019
		(0.000712)	(0.007909)		(0.004399)*	(0.159010)***			
	2	-0.0083	0.1033	0.0275	0.0330	0.7232	0.3090	103	0.125
		(0.007005)	(0.087115)		(0.016482)*	(0.179436)***			
Roddick Hewitt	1	0.0007	-0.0136	-0.0028	-0.0043	0.4546	0.2002	230	-0.031
		(0.002540)	(0.022533)		(0.007821)	(0.124565)***			
	2	-0.0007	0.0066	-0.0041	-0.0036	0.3441	0.1587	230	0.019
		(0.002487)	(0.027421)		(0.005886)	(0.139641)*			
Djokovic Hass	1	-0.0043	-0.0045	-0.0061	-0.0248	0.5250	0.2302	164	-0.009
		(0.003422)	(0.024291)		(0.007767)**	(0.092921)***			
	2	0.0044	0.0098	-0.0057	-0.0151	0.9123	0.4236	164	0.011
		(0.003168)	(0.037739)		(0.009268)	(0.171670)***			
Murray Ferrero	1	-0.0037	-0.0029	-0.0099	0.0156	0.3934	0.2574	102	-0.007
		(0.005323)	(0.012608)		(0.007307)*	(0.121573)**			
	2	-0.0011	0.0100	-0.0087	-0.0045	0.6403	0.3007	102	0.015
		(0.001403)	(0.027589)		(0.003991)	(0.145423)***			

Table 2.7. Price Discovery: Results (60s intervals). This table describes the results of the estimation of Equations (2.3) and (2.4). These equations examine the contribution to price discovery of the win market and the exotic set market. For this analysis, pricing samples were taken at intervals of 60 seconds for the full duration of each match. Newey-West heteroskedasticity and autocorrelation -consistent standard errors are in parentheses and .,*,**,*** indicates significance at the 10%, 5%, 1% and 0.1% levels respectively. The price discovery contribution of the exotic set market, as defined in Equation (2.6), is presented in the final column.

2.5 Conclusion

A recent scandal surrounding alleged 'spot-fixing' in cricket has focused negative attention on the more extreme exotic bets available. Nevertheless, it could be argued that exotic bets - which allow for wagering on outcomes other than the winner of a sporting event - allow the informed to capitalise on their precise, legally obtained, information. As information is revealed through betting, this may aid price discovery. On the other hand, if the informed were to disproportionately assemble in the exotic market, liquidity would decrease as those providing liquidity seek to limit their losses. If this illiquidity subsequently drives the informed out, this would leave the exotic market without any practical purpose.

Using in-play pricing data from the Men's Wimbledon Tennis Championship in 2009, we found that the price discovery contribution of the exotic set betting market was negligible. This suggests that providing large numbers of miniscule events for punters to bet on will have little impact on our ability to forecast events. Only when a player winning was a near inevitability - and the win market nearly became obsolete - did the exotic market contribute the majority of price discovery.

Our results have implications for the prediction markets literature. A prediction market facilitates wagering on the probability of an event, much like a betting market, but also 'produce[s] information externalities that can inform business and policy decisions'.[7] As markets are relatively efficient in aggregating dispersed information, studies have found that our ability to forecast events is aided by a prediction market (see Wolfers and Zitzewitz (2004) and Snowberg *et al* (2008) for surveys of the literature). It would be natural, therefore, to assume that increasing the number of markets would increase our overall forecasting ability. Our results suggest however, that at least where there are monetary payoffs, liquidity is critical in determining the price discovery contribution of the new market.

[7] Snowberg *et al* (2008)

References

- Ali, M., M., (1979). Some Evidence of the Efficiency of a Speculative Market. *Econometrica*, 47, 387-392.
- Asch, P., Quandt, R., E., (1987). Efficiency and Profitability in Exotic Bets. *Economica*, 54, 289-298.
- Biais, B., Glosten, L., Spatt, C., (2005). Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications. *Journal of Financial Markets*, 8, 217-264.
- Black, F., (1975). Fact and Fantasy in the Use of Options. *Financial Analysts Journal*, 31, 36-41+61-72.
- Chakravarty, S., Gulen, H., Mayhew, S., (2004). Informed Trading in Stock and Option Markets. *Journal of Finance*, **59**, 1235-1257.
- Chordia, T., Goyal, A., Sadka, G., Sadka, R., Shivakumar, L., (2006). Liquidity and the Post-Earnings-Announcement Drift. London Business School Working Paper.
- Chordia, T., Roll, R., Subrahmanyam, A., (2008). Liquidity and Market Efficiency. *Journal of Financial Economics*, 87, 249-268.
- Copeland, T., Galai, D., (1983). Information Effects and the Bid-Ask Spread. *Journal of Finance*, 89, 287-305.
- Dupont, D., (2000). Market Making, Prices, and Quantity Limits. *Review of Financial Studies*, 13, 1129-1151.
- Easley, D., O'Hara, M., Srinivas, P., S., (1998). Option Volume and Stock Prices: Evidence on Where Informed Traders Trade. *Journal of Finance*, **53**, 431-465.
- Easton, S., Uylangco, K., (2009). Forecasting Outcomes in Tennis Matches Using Within-Match Betting Markets. *International Journal of Forecasting*, 26, 564-575.

- Forrest, D., McHale, I., (2007). Anyone for Tennis (Betting)? *European Journal of Finance*, 13, 751-768.
- Garbade, K., D., Silber, W., L., (1983). Price Movements and Price Discovery in Futures and Cash Markets. *Review of Economics and Statistics*, 65, 289-297.
- 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.
- Gromb, D., Vayanos, D., (2010). Limits of Arbitrage: The State of the Theory. Paul Woolley Centre Working Paper.
- Hasbrouck, J., (1995). One Security, Many Markets: Determining the Contributions to Price Discovery. *Journal of Finance*, 50, 1175-1199.
- Hasbrouck, J., (2007). *Empirical Market Microstructure. The Institutions, Economics, and Econometrics of Securities Trading.* Oxford University Press.
- Hausch, D., B., Lo, V., S., Y., Ziemba, W., T., (2008). *Efficiency of Racetrack Betting Markets*, World Scientific.
- Hausch, D., B., Ziemba, W., T., (2008). *Handbook of Sports and Lottery Markets*, North Holland.
- Madhavan, A., (2000). Market Microstructure: A Survey. *Journal of Financial Markets*, 3, 205-258.
- O'Hara, M., (1995). *Market Microstructure Theory*. Blackwell Publishing.
- 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.
- Smith, M., A., Paton, D., Vaughan Williams, L., (2006). Market Efficiency in Person-to-Person Betting. *Economica*, 73, 673-689.

- Snowberg, E., Wolfers, J., Zitzewitz, E., (2008). Prediction Markets: From Politics to Business (and Back). In Hausch, D., B., Ziemba, W., T., (2008). *Handbook of Sports and Lottery Markets*, North Holland, 385-402.

- Spann, M., Skiera, B., (2009). Sports Forecasting: A Comparison of the Forecasting Accuracy of Prediction Markets, Betting Odds and Tipsters. *Journal of Forecasting*, 28, 55-72.

- Tetlock, P., C., (2004). How Efficient are Information Markets? Evidence from an Online Exchange. Harvard University Working Paper.

- Tetlock, P., C., (2008). Liquidity and Prediction Market Efficiency. Yale University Working Paper.

- Vaughan Williams, L., (2005), *Information Efficiency in Financial and Betting Markets*, Cambridge University Press.

- Wolfers, J., Zitzewitz, E., (2004). Prediction Markets. *Journal of Economic Perspectives*, 18, 107-126.

- Zitzewitz, E., (2006). Price Discovery Among the Punters: Using New Financial Betting Markets to Predict Intraday Volatility. Stanford University Working Paper.

Chapter 3

Evidence of In-Play Insider Trading on a U.K. Betting Exchange

Abstract

An open question in market microstructure is whether 'informed' traders have an advantage due to access to private, inside, information; or due to a superior ability to process public information. In this chapter we attempt to answer this question with data from a sports betting exchange taken during play. Uniquely, this allows us to time-stamp information events to the nearest second, and to ensure we are observing all relevant information regarding the value of an asset. We find evidence of inside information but not of a superior ability to process public information. The first finding suggests that a subset of the betting population are observing the action before the wider public (possibly due to delays in the television signal), and betting using this informational advantage.

3.1 Introduction

There is a rich history of empirical betting market studies to explore financial market theories. In particular, betting markets have proved a popular setting for tests of market efficiency, and interest has focused on a persistent anomaly: the favourite/long-shot bias, where returns on favourites exceed those on long-shots.[1] In this chapter we use betting market data to test two hypotheses from market microstructure.

The initial motivation for market microstructure research was a realisation that Walrasian equilibrium was a poor characterisation of the type of trade carried out on major stock exchanges.[2] In a continuous time trading environment an intermediary is either contracted to, or can extract profits from, the provision of liquidity. The costs of this intermediation distort the quoted bid and ask prices from the asset's fundamental value. Analysis of the formation of price, which had previously focused on the fundamentals of the asset, now needed to include the payoffs and preferences of the intermediary.

In the Glosten and Milgrom (1985) model, the specialist market-maker trades an asset with a population composed of informed traders, who have private, inside, information on the fundamental value of the asset, and liquidity traders, who trade randomly. A bid-ask spread is charged in order to offset losses to the informed with gains from liquidity traders. This is the adverse selection component of the bid-ask spread. Extrapolating this result, if a subset of the trading population has private information on the contents of a forthcoming public announcement, then the bid-ask spread will increase prior to, and during, the announcement.

Kim and Verrecchia (1994) endogenise the acquisition of private information by adding a trading group which processes information. Their information advantage ma-

[1] The first observation of the bias, in Griffith (1949), predated the market efficiency literature. Ottaviani and Sørensen (2008) survey the technical explanations of the bias. For a more general survey of market efficiency in betting markets see Vaughan Williams (2005).

[2] Garman (1976) was the author to coin the term 'market microstructure'. For a review of the market microstructure literature see O'Hara (1995), Madhavan (2000), Biais, Glosten and Spatt (2005) and, for empirical work, Hasbrouck (2007).

terialises after information on the asset is publicly announced, as they are able to create private information via their analysis. The specialist market-maker will therefore increase the bid-ask spread after a public announcement to offset losses to the information processors. This model interestingly proposes that information disclosure can increase, rather than limit, adverse selection if the signal received from information is noisy.

In the context of information arrival, the two models lead to the following non-competing hypotheses:

Hypothesis 1. The adverse selection component of the bid-ask spread arises due to asymmetric information prior to, and during, public information arrival.

Hypothesis 2. The adverse selection component of the bid-ask spread arises due to asymmetric abilities to process symmetric information after public information arrival.

A number of authors have examined the bid-ask spread in financial markets around significant information events in order to distinguish between the two hypotheses. Lee, Mucklow and Ready (1993) find that spreads widen both before and after earnings announcements, although the effect after an announcement is short-lived. Krinsky and Lee (1996) decompose the bid-ask spread into its adverse selection, transaction cost, and inventory-control components, and find that the adverse selection component of the bid-ask spread increases both prior to and after an earnings announcement. Gajewski (1999), in a study of earnings announcements on the Bourse de Paris, finds greater support for *Hypothesis 2*.

In our study we examine bid-ask spreads on a betting exchange, both before a sporting event, when information arrival is infrequent, and during an event, when information arrival is highly frequent. The advantage of such a setting is that information arrival can be time-stamped to the nearest second. Previously, information arrival in this context was identified only to the nearest minute (see Gajewski (1999)). In addition, while the announcement of information in a financial market does not preclude the prior existence, or subsequent arrival, of private information not observed by the public, in a sporting environment, once play has begun, all information relevant to the value of a betting asset is observable. This allows us to correctly isolate the effects of information arrival. A final advantage lies in the fact that betting market assets have a reasonable probability of default. The importance of new information for the value of a betting asset is greater than new information in a financial market, where the probabilility of default, regardless of the information content, is minimal. Adverse selection in a betting market should therefore be more pronounced.

The betting price data we use in this chapter is taken from a tennis match. Tennis is chosen because information events are of sufficient length and can be foreseen. For example, a tie-break is often important in determining the outcome of a match and takes place over a sufficient period of time for bettors to realise the importance of the play whilst it occurs.[3] In contrast, a goal is important in determining the outcome of a football match but is short-lived and cannot necessarily be foreseen. In our study, information periods are separated into four, with a *low information period* prior to play and during rain breaks; an *intermediate information period* whilst the match is in play; a *high information period* during tie-breaks; and a *post-high information period* immediately following a tie-break. We find that the adverse selection component of the bid-ask spread increases during our high information period, but decreases, even relative to our intermediate information period, immediately after the high information period. This lends support to *Hypothesis 1*, but not *Hypothesis 2*.

This brings us to the question of the nature of inside information on a betting market. Prior to play this could take the form of information on the wellbeing, fitness or determination of the athlete. Once the match has begun, such information is typically revealed to the public in the early stages of play. However, we observe that inside information increases during a match, and, further, appears to spike during moments of importance, such as tie-breaks. In other words, inside information is being created during a match. We therefore propose that informed bettors derive their advantage

[3]A tie-break is played at the end of a set if the players are tied on 6 games each. The winner of the tie-break wins the set.

from observing the action before the rest of the public. Television pictures typically transmit with a few seconds delay and therefore bettors with a faster transmission, or present at the game, are able to trade on an informational advantage.[4] To capitalise on this fleeting advantage, bettors would need to feed this information into a computer and initiate bets via an algorithm. In these circumstances, a high frequency trading strategy (trading after each point, for example) would generate significant returns.

The chapter is set out as follows. In Section 3.2 we describe the data and present some descriptive statistics. In Section 3.3 we outline the methodology and provide results. Section 3.4 concludes.

3.2 Data

The focus of our betting study is the Men's 2008 Wimbledon Tennis Final, between Roger Federer and Rafael Nadal. Described afterwards as 'one of the greatest finals of a grand-slam tournament' (*The Times*, July 8th 2008), this match attracted a lot of attention as it pitted the number one ranked player in the world (Federer) against the number two (Nadal), in the most prestigious of the grand-slam events. As an indication of the betting interest, Betfair, a betting exchange, matched GBP 28,334,894 on Nadal to win and GBP 20,802,434 on Federer to win.

We obtained Betfair betting price data for this match from Fracsoft, a company contracted to market historical pricing data for Betfair. The data available includes the quoted odds and respective volumes for each player to win, for 211 minutes and 48 seconds before the match begins, and also the 400 minutes and 44 seconds from the beginning of the match until the end. This gives us 36752 seconds of pricing data for this match. The match itself lasted for 4 hours and 48 minutes but there were two rain delays, during which betting could continue. Bets on an exchange can be traded in the form of a back bet (where the bettor receives the stake plus the odds if the event occurs), or a lay bet (where a bettor receives the stake if the event does not occur but is

[4] As we discuss in section 3.3, the betting exchange in question does take steps to limit this advantage.

liable for the odds if it does occur). From these quoted odds we calculated the back-lay (bid-ask) spread.

As well as a bet on a player to win, bets were traded on the score, in sets, by which a player would win. We used this data to identify the timing of tie-breaks. After a set ends (after a tie-break, for example) and a particular set score is no longer possible, bets cease to be traded. This allowed us to time-stamp the end of the tie-break. We then calculated the length of the tie-break with video footage of the match. As a result we were able to identify the *low information periods* (prior to play or during a rain break), the *intermediate information periods* (during play including tie-breaks), the *high information periods* (during a tie-break), and finally the *post-high information periods* (immediately following a tie break).[5]

As well as this match we carried out similar analysis on three matches at an earlier stage of this tournament. In these cases we did not have video footage, and so estimated the duration of tie-breaks as 5 minutes. Our statistical results for these matches were similar to those we report here, but because we were not able to cross-check the data with video footage, we limit the results we report to the 2008 Final.

This best of five set match finished 3 sets to 2 to Nadal. The 3rd and 4th sets both went to a tie-break, with Federer winning both to stay in the match. The *high information period* consists of the 3rd set tie-break, which lasted 456 seconds, and the 4th set tie-break, which lasted 816 seconds. The *post-high information period* is defined in our study as the 5 minutes (300 seconds) that immediately follow each of those tie-breaks.[6]

The criteria of a relevant betting asset for our study is as follows. The asset must have been traded during at least two periods which qualify as *high information periods* and two periods which qualify as *post-high information periods*. By this criteria, we

[5]There are undoubtedly times, other than during a tie-break, which could be classified as *high information periods*. We could also include break-points, set-points and match points. However, these periods are shorter than tie-breaks which may not give the liquidity providers sufficient time to react to their existence.

[6]We also classified the *post-high information period* as 1 minute following the tie-break and the results were unaffected.

analysed 4 bets: Federer to win, Nadal to win, Nadal to win 3-2, and Federer to win 3-2. All other set betting outcomes fell short of that criteria.

Table 3.1 outlines the descriptive statistics for the spreads quoted on the four assets. Spreads are converted into implied probability form. For example, if a back bet is quoted at 3-1, then the implied probability is $1/(3+1) = 0.25$. If the lay bet is quoted at 4-1, then the implied probability is $1/(4+1) = 0.2$, which results in a spread of $0.25 - 0.2 = 0.05$.

More generally, our measure of the bid-ask spread is

$$S = \frac{1}{O_b + 1} - \frac{1}{O_l + 1} \tag{3.1}$$

where O_b is the best back odds offered and O_l is the best lay odds offered.

The average spread is higher, and displays a greater standard deviation, in the *high information period* for all four assets. The difference is understandably most pronounced in assets 3 and 4, the set betting assets, which have the greatest variance in payoff. The average spread in the *post-high information period* is lower than the average spread in the *intermediate information period* for all four assets. This appears to suggest that spreads increase around important information arrival, but fall quickly once information has been revealed. This lends support to the Glosten and Milgrom (1985) based *Hypothesis 1*.

As an illustration of the evolution of the bid-ask spread from before play to during play, consider Figure 3.1. This figure represents the bid-ask spread in asset 1. The match begins at time 12708 and the bid-ask spread is on average higher and more volatile after this time. In the next section we present our empirical model.

Information period	1	2	3	4
All	0.0053 (0.0043)	0.0052 (0.0057)	0.0189 (0.0440)	0.0201 (0.0650)
Low	0.0038 (0.0000)	0.0032 (0.0000)	0.0053 (0.0011)	0.0055 (0.0013)
Intermediate	0.0061 (0.0051)	0.0062 (0.0068)	0.0261 (0.0530)	0.0278 (0.0793)
High	0.0098 (0.0084)	0.0131 (0.0168)	0.0910 (0.1636)	0.1170 (0.2200)
Post-High	0.0052 (0.0032)	0.0046 (0.0033)	0.0243 (0.0235)	0.0233 (0.0297)

Table 3.1. The mean bid-ask spread (Equation 3.1, to 4 d.p.) for asset 1 (Nadal to win), 2 (Federer to win), 3 (Nadal to win 3-2) and 4 (Federer to win 3-2) as measured in all, low, intermediate, high and post-high information periods. Standard deviations are in parentheses.

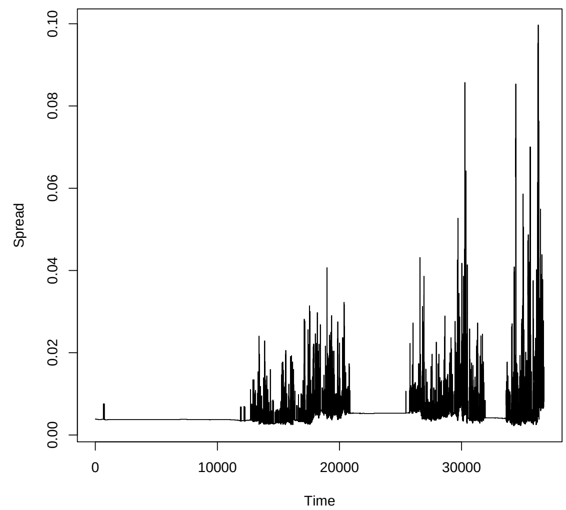

Figure 3.1: The bid-ask spread, as defined in Equation (3.1), for asset 1 (Nadal to win). The match began at Time=12708.

3.3 Methodology and Results

Our aim in this section is to control for the elements of the bid-ask spread which could be ascribed to transaction costs, inventory-control effects, or a lack of competition in liquidity provision, and therefore isolate the adverse selection component. In addition, we need to control for autocorrelation in the spreads as the spreads at nearby time periods are not independent.

To test our two hypotheses we considered the following regression model for each of the four assets.

$$S_t = \beta_0 + \beta_1 S_{t-1} + \beta_2 V_t + \beta_3 D_t + \beta_4 B_t + \beta_5 P_t + \epsilon_t \quad (3.2)$$

At time t, S_t is the spread as defined in equation (3.1), S_{t-1} is the spread at the previous time point, V_t is the sum of the volume available at the best three back and lay odds, and D_t, B_t and P_t are indicator variables, determining if t is during a *intermediate period* (during the match), a *high information period* (during a tie-break) or a *post-high information period* (in the 300 seconds following a tie break) respectively. $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are fixed coefficients and ϵ_t is an error term.

We include the spread at the previous time point to control for temporal dependencies between spreads, and a measure of volume to control for illiquidity which can cause spreads to widen irrespective of adverse selection. We also expect that our measure of volume controls for inventory-control effects, as the impact of a liquidity provider's inventory considerations would be limited in a liquid market.[7]

In this model β_0 has an economic interpretation. This is the component of the bid-ask spread which arises regardless of the informational considerations. This is a positive fixed transaction cost to cover the labour and computing resources involved in providing liquidity. We also require that $\beta_2 < 0$ as the bid-ask spread increases as competition, proxied by volume, decreases.

By *Hypothesis 1*, we expect that $\beta_3 > 0$ and $\beta_4 > 0$. When information is arriving, the liquidity providers should increase the bid-ask spread to guard against those with prior access to this information. For *Hypothesis 2*, we require that $\beta_5 > 0$. That is, the bid-ask spread is increased immediately after an important information event as a subset of the betting population has superior abilities to process this information and assess the fundamental value of the traded bet. In addition, *Hypothesis 2* requires $\beta_3 > 0$ as those with superior analytical abilities can put them to work on smaller information events.

Table 3.2 reports our results for the estimation of Equation (3.2). Diagnostic plots suggest the model assumptions are valid. In what follows, 'significant' indicates significance at the 5% level and 'highly significant' indicates significance at the 1% level. $\beta_0 > 0$ for all four assets and highly significant in three of them, and $\beta_2 < 0$ for all four assets and highly significant in three of them. As our rationale outlined above, there is evidence of a fixed transaction cost component of the bid-ask spread, and that the bid-ask spread decreases as volume increases.

[7] The inventory-control literature largely assumes that liquidity is provided by a contracted specialist who is obligated to provide liquidity continuously (see Stoll (1978), Amihud and Mendelson (1980) and O'Hara and Oldfield (1986)). When the liquidity provider does not have an obligation to trade continuously, as would be the case with a betting limit order trader, they can maintain a desired exposure to the event by placing market orders. In addition, the effect of inventory on price is transient and, particularly in a liquid market, should therefore be negligible.

We find that $\beta_3 > 0$, and significant, for all four assets. This concurs with our assumption that the arrival of information creates an adverse selection problem for those providing liquidity. Whether that adverse selection problem is due to inside information or the ability of certain bettors to process public information is answered by β_4 and β_5. We find that β_4 is positive and significant for all four assets. In other words, liquidity providers increase the bid-ask spread during significant information events due to asymmetric information at these times. This lends support to *Hypothesis 1*.

On the other hand, we find little support for *Hypothesis 2*. $\beta_5 < 0$ for all four assets and highly significant for two of them. One possible reason for the negative coefficient is that adverse selection is present in the periods after information arrival, but is simply less pronounced than adverse selection during information events. However, we have controlled for temporal dependencies between spreads by regressing on the spread in the last time period, and used Newey-West standard errors to account for serial correlation in spreads. We can be confident, therefore, that the adverse selection component of the bid-ask spread decreases after a significant information event. However, despite these efforts, we cannot formally reject *Hypothesis 2*. It is possible that bettors process information during tie-breaks (e.g. the implications of the tie-break score changing to 5-2 rather than 4-3), and that this is a contributing factor to a positive β_3. However, it is our proposition that the period immediately following a tie-break is where the opportunity for information processing is most prominent. As adverse selection is lower here than during the match as a whole, we can say that we have little evidence to support *Hypothesis 2*.

	1	2	3	4
β_0	0.001117	0.001318	0.000438	0.001084
	(7.82E-05)***	(0.000128)***	(0.000195)*	(0.000286)***
β_1	0.723595	0.769087	0.930341	0.935464
	(0.020057)***	(0.022484)***	(0.029465)***	(0.020452)***
β_2	-7.11E-10	-3.13E-09	-1.08E-08	-1.54E-07
	(1.18E-10)***	(3.38E-10)***	(9.74E-09)	(3.90E-08)***
β_3	0.00058	0.000151	0.001131	0.000564
	(3.97E-05)***	(2.63E-05)***	(0.000469)*	(0.000174)**
β_4	0.001067	0.001584	0.004809	0.006035
	(0.000181)***	(0.000299)***	(0.001713)**	(0.002399)*
β_5	-0.000275	-0.000282	-8.34E-05	-3.46E-05
	(6.50E-05)***	(9.18E-05)**	(0.000274)	(0.000428)
$Adj.R^2$	0.571369	0.645658	0.883073	0.887566
N	36762	36762	36762	36762

Table 3.2. The results of the estimation of Equation 3.2 for assets 1 to 4. Newey-West autocorrelation and heteroskedasticity consistent standard errors are in parentheses and *,**, and *** signals significance at the 5%, 1% and 0.1% levels respectively.

If bettors are trading on the basis of inside information, then the question remains as to the source of such information. Inside information in a betting market has traditionally been related to the fitness and determination of the athletes. Once the match has begun, such information is typically revealed to the public in the early stages of play. If this type of inside information was carrying over into the match, then we would expect little variation in the adverse selection component of the bid-ask spread both before and during a match. However, we observe, via our β_3 and β_4 coefficients, that adverse selection increases during a match, and, further, appears to spike during moments of importance, such as tie-breaks. In other words, inside information is being created during a match, and we propose that this inside information accrues to traders observing the action before the rest of the public.

There is reason to be believe that the pictures viewed on television are delayed with respect to viewing the action live. For television pictures to be transmitted, the images must be encoded, processed and then transmitted to the host broadcaster for further processing. They can then be transmitted to the home audience, or by satellite to other broadcasters around the world. Although information is not available on the delay in this particular instance, the delay may be significant even for those watching the event in the host country. If a bettor has a mobile device at the game, which would appear rather inconspicuous amongst the crowd, then information can be relayed to a computer instructed to trade algorithmically.

It must be acknowledged that our evidence for in-play insider trading is the fact that liquidity providers take mitigating actions to offset losses to such insiders. It may be argued that this evidence is circumstantial. Our approach, however, is in good company. In Shin (1993), the level of insider trading is inferred from the overround that a bookmaker charges on a series of horse races. The overround is the extent to which the sum of the implied win probabilities of all the horses exceeds 100%. This is the margin that the bookmaker claims. This margin, much like the bid-ask spread charged on a betting exchange, is the action that an intermediary takes to offset losses to those with private information.

Another issue is that Betfair does attempt to nullify the private information that accrues to those with a viewing advantage. For in-play markets there is a 1-5 second window (after a bet is matched) during which a liquidity provider can cancel the offer.[8] The length of the window depends on the company's estimation of the delay in television transmission. The aim of this window is precisely to deter trading on private information during a match. Our results, however, suggest that this window is either insufficient to remove the viewing advantage, or at least is perceived by those providing liquidity to be insufficient.

Our general results on the nature of informed trading differ slightly from those carried out on financial markets. Lee, Mucklow and Ready (1993), Krinsky and Lee (1996), and Gajewski (1999) find equal or greater support for *Hypothesis 2*, as for *Hypothesis 1.* A possible reason for this may lie in the different environment within which we test. There may be greater scope for detailed analysis of earnings announcements in a financial market than there is for the outcome of a tie-break in a tennis match. Although we would argue that a sophisticated bettor could calculate the conditional probability of a player winning, given the outcome of the last set, this does not correspond with our conversations with bookmakers. Although detailed statistical analysis is carried out

[8]http://help.betfair.com/contents/itemId/i65767339/index.en.html

prior to a match, in-play pricing is often determined by the bookmaker's 'feeling' on a game, whilst observing competitor's pricing to ensure that an arbitrage is not available. If a bookmaker believes there is little to be gained from detailed statistical analysis, it is unlikely that those providing liquidity on a betting exchange will feel the need to guard against other bettors using such analysis.

The task confronting bettors may just be simpler than that confronting a stock market trader. The odds quoted on a player to win are easily intrepreted as the implied probability of such an outcome. The efficient value of a stock, on the other hand, is the present value of all future returns, whether that be dividend payouts or capital gains. Analysing the effect of a company earnings announcement on such returns is therefore a complicated task.

A second possible explanation is the relative novelty of in-play betting exchanges. Although private information has been gathered, and sometimes created, on the outcome of sporting events for a long period, the possibility of trading in-play has only emerged in the last decade. As a result, returns from the possession of in-play inside information may be at an early and bountiful stage, if competition is low. Once the market develops, and opportunities diminish, it may be that informed bettors will follow financial market professionals and develop an alternative advantage via analytical techniques.

3.4 Conclusion

In this chapter we have set out to answer a fundamental question in market microstructure: whether 'informed' traders derive their advantage from access to inside information, or due to a superior ability to process public information. For this purpose we took data from a sports betting exchange during play. Uniquely, this allowed us to time-stamp information events to the nearest second, and to ensure we were observing all relevant information regarding the value of an asset. We found evidence of inside information but not of a superior ability to process public information. As traditional types of betting inside information (such as knowledge of the player's fitness) become stale once a match begins, our findings suggest that a subset of the betting population is creating its own inside information by observing the action before the wider public (possibly due to delays in the television signal), and betting using this informational advantage. To capitalise on this fleeting advantage, bettors would need to feed this information into a computer and initiate bets via an algorithm. In these circumstances, a high frequency trading strategy (trading after each point, for example) would generate significant returns.

In the U.K., the speed of television transmission differs substantially between those channels transmitted terrestrially, and those transmitted by satellite. The terrestrial transmission is noticeably faster. In 2008, the Wimbledon Final was televised on BBC1 which is available on terrestrial television. Most of the tennis played during the year however is only available on satellite television. This creates the possibility that the effect we have observed here may be more pronounced elsewhere.

References

- Amihud, Y., Mendelson, H., (1980), Dealership Market. Market-Making with Inventory, *Journal of Financial Economics*, **8**, 31-53.

- Biais, B., Glosten, L., Spatt, C., (2005), Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications, *Journal of Financial Markets*, **8**, 217-264.

- Gajewski, J., F., (1999), Earnings Announcements, Asymmetric Information, Trades and Quotes, *European Financial Management*, **5**, 411-423.

- Garman, M., B., (1976), Market Microstructure, *Journal of Financial Economics*, **3**, 257-275.

- 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-Racing Bettors, *American Journal of Psychology*, **62**, 290-294.
- Hasbrouck, J., (2007), *Empirical Market Microstructure. The Institutions, Economics, and Econometrics of Securities Trading*, Oxford University Press.
- Kim, O., Verrecchia, R., E., (1994), Market Liquidity and Volume Around Earnings Announcements, *Journal of Accounting and Economics*, **17**, 41-67.
- Krinsky, I., Lee, J., (1996), Earnings Announcements and the Components of the Bid-Ask Spread, *Journal of Finance*, **51**, 1523-1535.
- Lee, C., M., C., Mucklow, B., Ready, M., J., (1993), Spreads, Depths, and the Impact of Earnings Information: An Intraday Analysis, *Review of Financial Studies*, **6**, 345-374.
- O'Hara, M., Oldfield, G., (1986), The Microeconomics of Market Making, *Journal of Financial and Quantitative Analysis*, **21**, 361-376.
- O'Hara, M., (1995), *Market Microstructure Theory*, Blackwell Publishing.
- 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.
- Madhavan, A., (2000), Market Microstructure: A Survey, *Journal of Financial Markets*, **3**, 205-258.
- Shin, H., S., (1993), Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims, *The Economic Journal*, **103**, 1141-1153.
- Stoll, H., R., (1978), The Supply of Dealer Services in Securities Markets, *Journal of Finance*, **33**, 1133-1151.
- Vaughan Williams, L., (2005), *Information Efficiency in Financial and Betting Markets*, Cambridge University Press.

Chapter 4

Examining Agency Conflict in Horse-Racing

Abstract

We study U.K. horse-racing for signs of conflict between horse owners (principals) and trainers (agents). Trainers often prepare their own horses for races, in addition to having outsiders' horses in their care. Utilising betting market data to infer the expected performance of a horse, we find that owner-trainer horses outperform outsider-trainer horses, indicating that this principal-agent relationship is characterised by agent shirking. If the owner holds a large proportion of the horses in the trainer's stable, the shirking effect is mitigated, but not eradicated. In a separate result, we find that outsider-trainer horses are more inconsistent than their owner-trainer peers. As inconsistency is a sign of betting market manipulation, this suggests that the agent in this setting extracts a second, informational, rent from the principal.

4.1 Introduction

We employ agents to undertake a number of frequent tasks on our behalf. Agents are entrusted to find us a job, manage our investments, and school our children. As the effort and performance of agents is imperfectly observed, this creates ample opportunities for conflict, and often welfare losses for the principal. For example, Rutherford *et al* (2005) and Levitt and Syverson (2008) find that estate agents sell their client's houses for less than their own houses, after controlling for observable house characteristics. Similarly, Ang *et al* (2000) find that managers who have little stake in a small firm spend more on expenses and run the firm less efficiently than firms with a single owner-manager. The economic study of agency conflict dates back at least as far as Smith (1776), who analysed conflict in sharecropping, where a landowner allows a tenant to farm their land in return for a cut of their crop. A more formal treatment of principal-agent issues is provided by Laffont and Martimont (2001).

In this chapter we examine agency conflict in horse-racing. Race-horse trainers are responsible for the welfare of horses in their stable, supervising how the animal is fed, how hard the horse is run prior to a race, and often deciding in which races the horse will run. Many trainers (agents) divide their time between the preparation of their own horses and those of outsiders (principals). While the trainer receives all of the win purse if their horse is successful, they receive only a fraction (typically 10%[1]) for training an outsider's horse to victory. In addition, and crucially for any study of agency conflict, much of the trainer's effort is unobservable; moreover the asymmetric information problem is exacerbated by there being an animal at the centre of the conflict. To illustrate the opacity of the trainer's actions, Fox (2005) likens the race-horse trainer to a tribal shaman, or witch-doctor, whose success is ascribed to impressive work, and whose failure is due to factors outside of their control.

There are a number of elements to horse-racing that make it an interesting arena

[1] Boyle *et al* (2007) indicates that 10% is the typical commission the trainer can earn but Scott (1968) suggests the commission can be as much as 50% if the owner wishes to align the trainer's incentives with their own.

for the study of agency conflict. Firstly, in contrast to the infrequent sale of a house, horses run on a regular basis over the course of their careers, yielding a large sample of performance data. We use data on every horse-race in the U.K. between 2005 and 2010. Secondly, the interaction of horse trainer and owner is dynamic, as the owner can withdraw a horse from a trainer's stable if performance is deemed unsatisfactory; the estate agent is unlikely to fret to the same degree over repeat business from a house-seller. Thirdly, there is no need to infer the intrinsic ability of a horse from its observable characteristics; there is a betting market from which to gauge the pre-race expected performance of a horse. Finally, the horse owner is relatively empowered compared to the owner of a small firm; whereas removing a firm manager can be problematic and time-consuming, the horse can often be removed from a trainer's stable on short notice and without compensation.

Utilising betting market data to infer the expected performance of a horse, we find that owner-trainer horses outperform when compared with outsider-trainer horses. This suggests that this particular agency conflict is characterised by the agent shirking. Alternatively, it is plausible that trainers - who undoubtedly possess private information on the racing game - are better able to identify the good horses unappreciated by the betting public. To control for this, we examine the performance of older, higher-class, horses for which the private information set should be smaller. As in the broader study, owner-trainer horses outperform outsider-trainer horses in this subset of animals.

One way in which an owner can mitigate the shirking effect is to entrust the preparation of a number of horses to the same trainer. This way the trainer should be careful not to disappoint the owner for fear of losing a substantial proportion of their stable and livelihood. We calculate the dominance of a trainer's stable by each owner, and find that owning a large proportion of the horses under a trainer's care can mitigate, but not eradicate, the shirking effect.

In horse-racing there is potentially a second source of income besides the prize money. If trainers, or owners, have private information on the upcoming performance of a horse, they can bet for profit; indeed, in the U.K. it is legal for them to do so. Trainers have the potential to create such private information, as they control the preparation of horses in their care. One sign of manipulation of horse performance for betting market gain is inconsistency; initiating a poor performance by the horse one week - thereby causing the betting public to downgrade their estimate of the horse's ability - only for the horse to outperform relative to the betting market's downgraded expectations the next week, when in fact the horse was simply performing at its true level in the second race.

We examine whether such performance manipulation is a second, informational, source of agency conflict. After controlling for other factors that affect the consistency of performance, we find that horses trained by an outsider are more inconsistent than horses trained by their owner. This suggests that trainers not only exert less effort when training an outsider's horse, but also, on average, exploit these horses for betting market gain. These results give empirical weight to anecdotal evidence that horse-racing is at least partially manipulated, and that the trainer - with their privileged access to information - is front and centre in this manipulation.[2]

The evidence of performance manipulation in horse-racing also corresponds with evidence of betting market-induced corruption in other sports. For example, Wolfers (2006) finds evidence of point-shaving - where a heavily-favoured team wins but fails to beat the bookmaker's spread - in college and professional basketball. Borghesi (2008) and Bernhardt and Heston (2010) suggest that such indirect studies of corruption, which rely on subtle inferences, should test an alternative hypothesis to ensure the robustness of the result.

In our case, an alternative explanation for the greater consistency of owner-trainer horses is that the trainer - with their superior knowledge of racing - is able to select the most consistent horses. If the trainer does have private information, it is much more likely to be related to younger, lower-profile, horses. This explanation can be discounted, however, because the inconsistency of outsider-trainer horses is apparent even after controlling for the class in which a horse runs, and its age.

[2] Scott (1968), in an anthropological study of a U.S. racetrack, details how widespread performance manipulation is, and the central role trainers play in this due to their informational advantage over other actors.

A recent paper by Boyle *et al* (2007) examines horse-racing results data from New Zealand to search for conflict between horse owners and trainers. On average, they find that the performance of horses is unaffected by the owner trainer relationship, though horses trained in small stables do show some signs of agency-induced underperformance. Their study differs from ours in that betting market data is not used to infer a horse's expected performance, nor do the authors consider the possibility that the trainer's manipulation of the performance of outsiders' horses - and the resultant opportunity for profitable betting - may be a second source of agency conflict.

The rest of this chapter is structured as follows. Section 5.2 describes the data set, and how the expected performance of a horse is inferred. Section 5.3 contains the empirical analysis of horse performance and consistency, and Section 5.4 concludes.

4.2 Data

We collected data on every horse race run in the U.K. between 1st January 2005 and 31st December 2010 inclusive, from Betwise, a betting information company. After discarding data on races without a designated class - which is an important control variable for our later analysis - we were left with 559,383 horse-level performance observations. The race data includes the date of the meeting, the type of race (e.g. flat, national hunt), the distance over which the race is run, the handicap system, the class of the race (which ranges from 1 to 7), and the win prize money (which ranges from GBP 0 to GBP 1,000,000). Summary statistics on horse and race characteristics can be found in Table 5.1.

All (N=559383)	Mean	Median	Max	Min	Std. Dev	Skewness	Kurtosis
Age	4.95	4	22	2	2.33	0.84	3.28
Starting Price (Odds)	22.6	13	501	1.02	29.2	3.64	24.32
Class	4.41	5	7	1	1.41	-0.71	3.09
Distance (000's of yards)	2.72	2.2	7.92	1.1	1.4	0.66	2.27
No. of Horses in Race	11.53	11	40	2	3.74	0.89	6.16
Win Prize Money (000's of GBP)	12.65	5.24	1000	0	38	13.83	275.4
$Owner = Trainer$ (N=65861)	Mean	Median	Max	Min	Std. Dev	Skewness	Kurtosis
Age	6	6	16	2	2.61	0.54	2.72
Class	4.78	5	7	1	1.19	-0.67	3.69
Distance (000's of yards)	3.1	3.52	7.92	1.1	1.47	0.18	1.79
No. of Horses	11.49	12	40	2	3.39	0.44	4.33
Win Prize Money (000's of GBP)	7.11	4.5	1000	0	18.14	25.03	1007.46
Starting Price (Odds)	30.7	17	501	1.17	37.27	2.99	16.85
$Owner \neq Trainer$ (N=493522)	Mean	Median	Max	Min	Std. Dev	Skewness	Kurtosis
Age	4.8	4	22	2	2.25	0.87	3.34
Class	4.36	5	7	1	1.42	-0.69	2.99
Distance (000's of yards)	2.67	2.2	7.92	1.1	1.39	0.73	2.4
No. of Horses	11.54	11	40	2	3.78	0.93	6.29
Win Prize Money (000's of GBP)	13.39	5.44	1000	0	39.86	13.23	251.47
Starting Price (Odds)	21.52	12	501	1.02	27.78	3.74	25.74

Table 5.1. Summary statistics for horse and race characteristics for all U.K. races between 1st January 2005 and 31st December 2010. The full sample is described in the top panel, with the *owner = trainer* **sub-sample in the middle panel, and the** *owner ≠ trainer* **sub-sample in the bottom panel.**

For each horse we have their age, and the identity of their owner and trainer at the time of the race. There are 55,763 unique horses in our sample, with 27,861 unique owners and 2,944 unique trainers. To determine whether the owner was also the trainer of the horse, we took the surname of the trainer, then ran a regular expression to test whether any sequential part of the owner name was an exact match on the trainer surname. If it was, our *owner=trainer* dummy variable equals 1, and 0 otherwise. This algorithm captures, at least in part, the frequent occurrence of a trainer registering a horse they own in their wife's name.[3] By this method, 65,861 horse-level performances are given by horses where the owner is the trainer, with the remaining 493,522 performances given by horses where the owner is not the trainer. Summary statistics on

[3]To ensure that our results are not driven by common names (e.g. Smith, Brown), we also altered the algorithm so that only owners and trainers with an exact match (i.e. A Brown) were classified as one individual. Our results were unaffected.

these two sub-samples are also in Table 5.1. On average, owner-trainer horses are older, compete in higher class races (but for less prize money), and are quoted longer odds, than outsider-trainer horses. However, these differences are relatively small, and are taken into account in the multivariate analysis in Section 5.3.

We are also interested in the extent to which an owner dominates a trainer's stable. This is calculated by counting the number of horses in a trainer's stable in any year and defining our *owner dominance* variable as the proportion (between 0 and 1) of these horses which the owner in question holds. For example, suppose a trainer had 10 horses, of which 2 were his own and 8 belonged to one other individual. For the trainer's two horses we would have the *owner=trainer* dummy equal to 1, and the *owner dominance* variable equal to 0.2. For the remaining 8 horses, the *owner=trainer* dummy would be 0, with the *owner dominance* variable equal to 0.8. The expectation is that the outsider-owner can offset part of the agency costs by holding a large proportion of the remaining horses under the same trainer's supervision.

In order to infer the expected performance of each horse, we collected the Starting Price, which is a summary measure of the odds quoted by British bookmakers at the time the race began. From here we can infer a predicted finishing position in the race for each horse. For example, a horse with odds of 2-1 (GBP 2 returned for every GBP 1 staked, if the horse wins) can be expected to finish above a horse with odds of 4-1.[4]

We then compared the predicted finishing position of each horse with their actual finishing position. All horses which failed to finish a race were classified as finishing last. We then used the actual and predicted finishing positions, and number of horses in the race, to construct a measure of horse performance:

$$RelativePerformance = \frac{PredictedFinish - ActualFinish}{NumberofHorses} \tag{4.1}$$

[4]There is the possibility of errors in this estimation of predicted finishing position. For example, consider a temperamental horse, who will either win or fail to finish. This horse may be expected to finish behind a slower, more consistent, horse but the win odds will not capture this. Nevertheless, in the absence of place (top 2/3 finish) odds, the win odds are a good proxy for the predicted finishing position.

Relative performance has a mean of -0.0266, a median of 0, a minimum of -0.97 and a maximum of 0.96. The standard deviation is 0.3084, with skewness of -0.145 and kurtosis of 0.1980. The measure is negatively skewed because non-finishing horses are classified as having finished last. If there is more than one horse that fails to finish, the average *relative performance* is less than 0. If no more than one horse fails to finish, the average for that race is 0.

In order to infer the consistency of the horse, we are also interested in the absolute value of *relative performance*:

$$AbsoluteRelativePerformance = |\frac{PredictedFinish - ActualFinish}{NumberofHorses}| \tag{4.2}$$

This measure treats a horse that outperforms exactly the same as a horse that underperforms. Suppose a horse runs in a 8-horse race. The horse is tipped to finish 1st but instead finishes 5th. This leads to a *absolute relative performance* observation of 0.5. If another horse is predicted to finish 5th but actually finishes 1st, the *absolute relative performance* observation would also be equal to 0.5. The aim of this measure is to capture instances when the trainer, or owner, could have profited from the performance of the horse, either due to under-performance or out-performance. A bettor can indeed now profit from under-performance as betting exchanges (such as Betfair in the U.K.) facilitate betting against a horse. Further, the potential payoff to the bettor increases in the degree of under-performance/out-performance.

4.3 Empirical Analysis

In the first part of this section we are interested in the effect that the separation of ownership and management (training) in horse-racing has on performance. Further, we examine the extent to which an outsider-owner can mitigate the agency problem by placing a number of horses with the same trainer, thereby increasing the trainer's dependence on the owner.

In Table 5.2, we display the results of a regression of our *relative performance* measure on the *owner=trainer* dummy and the *owner dominance* variable for the full 6 year sample. The intercept term is negative, due to the negative skew in *relative performance*. More importantly, we observe that owner-trainer horses outperform outsider-trainer horses at the 0.1% significance level. As the average race is between 11.53 horses (see Table 5.1), the coefficient in regression 1, Table 5.2 implies that owner-trainer horses, on average, outperform outsider-trainer horses by 0.35 places. This agency-induced problem can be mitigated by an owner's dominance of a trainer's stable, as this coefficient is positive and significant at the 0.1% level. However, examining the coefficients of the model, we can see that an owner who completely dominated an outsider-trainer's stable could not expect to completely avoid agency costs in the form of underperformance.

We ran supplementary regressions using fixed effects - also displayed in Table 5.2 - estimated using the White method to allow for clustering of errors at the handicap-type, horse, trainer, owner, race-type, and race level. The fixed effects for handicap-type, horse, race-type, and race have little effect on the remaining coefficients. On the other hand, the trainer fixed effect, captured in Regression 4, serves to reduce, but not eliminate, the coefficient associated with the *owner=trainer* dummy, and reverse the coefficient associated with *owner dominance*. We conjecture that this implies that most of the agent shirking is concentrated with a number of 'bad-egg' trainers. Similarly, the coefficients are reduced when owner fixed effects are incorporated, suggesting that a number of 'naive' owners are being disproportionately hit by agency costs. In race terms, owner-trainer horses outperform outsider-trainer horses by only 0.07 places once trainer fixed-effects are incorporated, and only 0.15 places when owner fixed-effects are incorporated.

Dep.Variable: Relative Performance

	1	2	3	4	5	6	7
Intercept	-0.032284	-0.033817	-0.032298	-0.02598	-0.029335	-0.033429	-0.034754
	(0.000478)***	(0.002876)***	(0.027893)	(0.002201)	(0.014377)*	(0.000546)***	(0.038014)
Owner=Trainer Dummy	0.030231	0.034705	0.035163	0.006956	0.012966	0.03283	0.038689
	(0.001432)***	(0.001682)***	(0.011492)**	(0.001642)***	(0.006780).	(0.001434)***	(0.010001)***
Owner Dominance	0.016303	0.024165	0.011872	-0.011473	0.009124	0.022859	0.027831
	(0.002120)***	(0.002137)***	(0.017179)	(0.004134)**	(0.010253)	(0.002125)***	(0.003555)***
Incl. Handicap Fixed Effects	No	Yes	No	No	No	No	No
Incl. Horse Fixed Effects	No	No	Yes	No	No	No	No
Incl. Trainer Fixed Effects	No	No	No	Yes	No	No	No
Incl. Owner Fixed Effects	No	No	No	No	Yes	No	No
Incl. Race-Type Fixed Effects	No	No	No	No	No	Yes	No
Incl. Race Fixed Effects	No	No	No	No	No	No	Yes
R^2	0.001474	0.00486	0.115379	0.033707	0.074864	0.003844	0.01327
No. of Observations	559383	559383	559383	559383	559383	559383	559383

Table 5.2. Coefficient estimates when relative performance was regressed on an owner-trainer dummy variable (equalling one when the owner is also the trainer) and owner dominance (between 0 and 1, equalling 1 when the horse owner holds all horses in a trainer's stable). Regression 1 is pooled, while regressions 2, 3, 4, 5, 6 and 7 incorporate handicap-type, horse, trainer, owner, race-type, and race fixed effects respectively. White heteroskedasticity-consistent standard errors - including allowances for clustering of errors in fixed effects - are in parentheses. *,**,*, and . indicates significance at the 0.1%, 1%, 5% and 10% levels respectively.**

One plausible explanation for the significance of the *owner=trainer* dummy is that trainers have private information on horses and therefore buy horses which will, on average, outperform relative to betting market expectations. To control for this we re-run the first regression, this time sampling only horses running in races of class 5 or over, and of age 5 or over. Older, higher-grade, horses are more likely to be well-known by the betting public, and therefore the amount of private information, on which the trainer can capitalise, is likely to be less. This choice of class and age removes around half of the sample in each case. As shown in Table 5.3, the *owner=trainer* dummy and the *owner dominance* variable are still positive and significant determinants of horse performance for these two subsamples at the 0.1% level.

Another concern may be that the focus on horse-level performances is dominated by horses than run often, and stables that contain a large number of horses. To a certain extent this is controlled for by the horse and trainer fixed effects, but in regression 4 of Table 5.3 we consider only horses that are part of a stable of 149 horses or less, and in regression 5 we consider only horses that ran in 18 races or less during our sample period. Again, both sub-samples comprise approximately half of the full sample size. The owner-trainer effect is still present in these two sub-samples - highlighting the robustness of this result - but the dominance of a trainer's stable, by an owner, is not a significant factor amongst horses that run less frequently. A possible explanation is that a horse which runs infrequently - perhaps because of injuries - is of less importance to both the trainer and the owner.

Dep.Variable: Relative Performance

	1	2	3	4	5
	All	Class ≥ 5	Age ≥ 5	Stable Size ≤ 149	No. of Runs ≤ 18
Intercept	-0.032284	-0.029741	-0.040987	-0.003882	-0.045664
	(0.000478)***	(0.000640)***	(0.000714)***	(0.000691)***	(0.000651)***
Owner=Trainer Dummy	0.030231	0.019395	0.027539	0.009799	0.038783
	(0.001432)***	(0.001788)***	(0.001944)***	(0.001763)***	(0.001951)***
Owner Dominance	0.016303	0.029409	0.035887	0.018473	0.000528
	(0.002120)***	(0.002895)***	(0.002829)***	(0.002564)***	(0.002735)
Incl. Handicap Fixed Effects	No	No	No	No	No
Incl. Horse Fixed Effects	No	No	No	No	No
Incl. Trainer Fixed Effects	No	No	No	No	No
Incl. Owner Fixed Effects	No	No	No	No	No
Incl. Race-Type Fixed Effects	No	No	No	No	No
R^2	0.001474	0.001351	0.002749	0.000633	0.001662
No. of Observations	559383	298191	277838	284561	290550

Table 5.3. Coefficient estimates when relative performance was regressed on an owner-trainer dummy variable (equalling one when the owner is also the trainer) and owner dominance (between 0 and 1, equalling 1 when the horse owner holds all horses in a trainer's stable). Regression 1 is pooled, regression 2 considers only horses when run in races of Class 5 or over, regression 3 considers only horses aged 5 and over, regression 4 considers only horses in a trainer's stable of less than or equal to 149 horses, and regression 5 considers only horses that ran 18 times or less over the sample period. White heteroskedasticity-consistent standard errors are in parentheses. *,**,*, and . indicates significance at the 0.1%, 1%, 5% and 10% levels respectively.**

Our second search for agency conflict focuses on the consistency of horses. A horse can be inconsistent due to age, the class of the race, the distance over which they run, the prize money on offer and also the number of horses in a race. In particular, we expect that horses running in higher class races, with greater prize money, should be more consistent. Boyle (2008) and Johnson *et al* (2010) find that high-class races finish closer to betting market expectations than low-grade races. This could be due to the greater inherent consistency of better horses, or due to the greater monetary incentives for maximum effort at the higher-grade races. These monetary incentives should also result in a negative relationship between *absolute relative performance* and win prize money. In addition, we believe that races run with a greater number of horses should be less predictable - yielding larger *absolute relative performance* measures - as there is greater potential for collisions and for a horse to be impeded. Finally, we are *a priori* unsure of the relationship between the distance of the race, and the age of the horse on *absolute relative performance.*

In an opaque environment where the horse's true ability is difficult to observe, inconsistency can also be a sign of betting market manipulation. A poor performance one week can lead the betting market to downgrade their estimate of the horse's ability, driving up the odds, and potential returns to betting, the next race the horse runs. We assess whether, after controlling for observable determinants of inconsistency, outsider-trainer horses are more inconsistent than owner-trainer horses. If so, this suggests the existence of a second, informational form of agency conflict.

We regress the *absolute relative performance* measure on the *owner=trainer* dummy and the *owner dominance* variable, with results displayed in Table 5.4. Reassuringly, we find that the control variables instituted have, in the main, the expected impact on *absolute relative performance*. Specifically, *absolute relative performance* is decreasing in win prize money (except in one specification) and the class of the race, and increasing in the number of horses in the race. In the focus of our study, we find that outsider-trainer horses are more inconsistent, as *absolute relative performance* is lower when the owner is also the trainer. This result is robust to the introduction handicap-type, horse, trainer and owner fixed effects. The introduction of trainer fixed effects reduces the coefficient associated with the owner-trainer dummy, again suggesting that any manipulation of horse performance is concentrated with a number of 'bad-egg' trainers.

It is less clear, however, that this conflict can be mitigated by the owner dominating a trainer's stable. The coefficient associated with this variable is not negative in all cases. It is perhaps surprising that outsider-owners can limit the shirking of trainers by placing a number of horses in the same stable, but appear less able to deter the manipulation of horse performance for betting purposes with the same policy.

A natural next step would be to develop a betting strategy that capitalises on the results outlined in this paper. This has not proved feasible for two reasons. Firstly, the over-round (the extent to which the sum of implied win probabilities quoted by a bookmaker exceeds 1) renders most strategies unprofitable, even if they increase returns above the blanket strategy of betting on all horses. A second reason - more closely associated with the betting market manipulation outlined in the paper - is that any trainer manipulating the performance of horses would, by necessity, attempt to disguise their actions. A horse may be deliberately under-performed for a number of races until the odds on a win have lengthened sufficiently for the trainer to push them for a win. Alternatively, the advent of betting exchanges - where bettors can wager on a horse to lose - could mean that the payoff is actually reaped when the horse underperforms. Therefore, while (unnecessary) inconsistency indicates betting market manipulation, it does not lead us to a profitable strategy.

There are two other issues that warrant discussion. The first is the role of the jockey; specifically whether the jockey would be in a position to manipulate horse races without the consent of the trainer. The jockey would be responsible for implementing the parts of performance manipulation that the trainer does not control (i.e. in the race itself), but it is unlikely that the jockey can manipulate horse performance without the trainer's knowledge. A jockey does not have a significant informational advantage over the trainer, who, after all, is in charge of all elements of the preparation of the horse for races.

A second issue is that, under certain circumstances, it may also be in the interests of trainers to manipulate the performance of their own horses. For this to occur, however, we would require the prize money to be very small compared to the potential betting income. As trainers receive only a percentage of any prize money, it must be optimal for the trainer to manipulate horse performance if it is optimal for the owner to do so, but it is not necessarily optimal for the owner to engage in manipulation if it optimal for the trainer to do so. Therefore, a trainer with both their own horses and those of outsiders in their care will first find it optimal to manipulate the performance of the outsiders' horses.

It should be noted that there are similarities between the trainer's situation - where the betting market provides insurance against poor horse performance - and that of a CEO who is paid in stocks and stock options but is allowed to hedge their exposure to the firm's cash flows. Bettis *et al* (2001) document that, in their sample, firm insiders use derivatives to reduce their stake in the firm by 25%, on average. A *New York Times* article (Eric Dash, Feb 5th 2001) describes such practices by Goldman Sachs staff. Theoretical analysis of this conundrum is provided by Gao (2010), who shows that firms will increase the sensitivity of CEO pay to performance (i.e. weight compensation further toward bonuses rather than base salary) to offset the effect that hedging has in reducing the CEO's exposure to firm value. Should data on the various incentive contracts in horse-racing become available, it would be instructive to identify whether the win commission paid to trainers is higher in high/low class races, or wherever the opportunity to hide hedging transactions (i.e. bet against the horse) is greater.

Another avenue for further work is to investigate the effect of career concerns on the performance of a trainer's horse. Gibbs and Murphy (1992) illustrate that older managers derive less motivation from implicit incentives such as career concerns, and therefore require greater explicit (pay) incentives in order to exert effort. The addition of data on the trainer's age may shed light on why, on average, implicit incentives are not sufficient to alleviate the agency problems described in this chapter. For example, the agency effects may be concentrated with older trainers whereas younger, career-minded trainers may be less problematic.

Dep. Var: Absolute Relative Performance

	1	2	3	4	5	6	7
Intercept	0.235279	0.201061	0.231978	0.192526	0.183706	0.184213	0.195775
	(0.000313)***	(0.001547)***	(0.005208)***	(0.043023)***	(0.003499)***	(0.015061)***	(0.001898)***
Owner=Trainer Dummy	-0.003716	-0.005640	-0.003405	-0.006724	-0.002702	-0.007342	-0.00402
	(0.000932)***	(0.000942)***	(0.001014)***	(0.002882)*	(0.001074)*	(0.003758).	(0.000939)***
Owner Dominance	0.001981	-0.004603	-0.003833	-0.003120	0.004728	0.00323	-0.004954
	(0.001398)	(0.001402)**	(0.001444)**	(0.002806)	(0.002211)*	(0.002949)	(0.001401)***
No. of Horses		0.002389	0.002224	0.002361	0.00244	0.002484	0.003052
		(7.56E-05)***	(9.14E-05)***	(0.000140)***	(7.73E-05)***	(8.66E-05)***	(7.81E-05)***
Distance (000's of yards)		0.002902	1.22E-03	0.008347	0.004543	0.005312	0.004376
		(0.000254)***	(0.000480)*	(0.002356)***	(0.000390)***	(0.000585)***	(0.000449)***
Class		-0.004147	-0.00339	-0.004734	-0.00278	-0.003148	-0.00376
		(0.000222)***	(0.000402)***	(0.000477)***	(0.000269)***	(0.000416)***	(0.000234)***
Win Prize Money (000's of GBP)		-0.000025	-1.66E-05	0.000033	-2.19E-05	-1.62E-05	-5.56E-05
		(8.39E-06)**	(1.47E-05)	(9.30E-06)***	(8.93E-06)**	(8.82E-06).	(8.42E-06)***
Age		0.003729	-0.00198	0.002886	0.004677	0.004514	0.002145
		(0.000144)***	(0.000742)**	(0.007210)	(0.000376)***	(0.001940)*	(0.000166)***
Incl. Handicap Fixed Effects	No	No	Yes	No	No	No	No
Incl. Horse Fixed Effects	No	No	No	Yes	No	No	No
Incl. Trainer Fixed Effects	No	No	No	No	Yes	No	No
Incl. Owner Fixed Effects	No	No	No	No	No	Yes	No
Incl. Race-Type Fixed Effects	No	No	No	No	No	No	Yes
R^2	0.000028	0.006617	0.024986	0.116652	0.015558	0.064459	0.009586
No. of Observations	559383	559383	559383	559383	559383	559383	559383

Table 5.4. Coefficient estimates when absolute relative performance (a measure of horse inconsistency) was regressed on an owner-trainer dummy variable (equalling one when the owner is also the trainer) and owner dominance (between 0 and 1, equalling 1 when the horse owner holds all horses in a trainer's stable). Regression 1 is pooled, regression 2 is pooled and includes control variables, while regressions 3, 4, 5, 6 and 7 incorporate handicap-type, horse, trainer, owner and race-type fixed effects respectively. White heteroskedasticity-consistent standard errors - including allowances for clustering of errors in fixed effects - are in parentheses. *,**,*, and . indicates significance at the 0.1%, 1%, 5% and 10% levels respectively.**

4.4 Conclusion

We examine horse-racing for signs of conflict between the principal - the horse owner - and the agent, the trainer entrusted with preparing the horse for racing. We find that owner-trainer horses outperform outsider-trainer horses, though this effect can, in some cases, be mitigated by the owner placing a number of horses with the same trainer. We also find a second agency cost, where outsider-trainer horses are more inconsistent than owner-trainer horses, suggesting that trainers, on average, use outsider's horses for betting market gain.

In race terms, the separation of ownership and management leads to, on average, under-performance of 0.35 places for a horse trained by an outsider. It is perhaps surprising that the betting market does not incorporate information on the ownership of horses and its implications for performance. It is due to this oversight, however, that we are able to identify the costs that owners bear when delegating management of their asset to an outsider.

References

- Ang, J., S., Cole, R., A., Wuh Lin, J., (2000). Agency Costs and Ownership Structure. *Journal of Finance*, 55, 81-106.

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

- Bettis, J., C., Bizjak, J., M., Lemmon, M., L., (2001). Managerial Ownership, Incentive Contracting, and the Use of Zero-Cost Collars and Equity Swaps by Corporate Insiders. *Journal of Financial and Quantitative Analysis*, 36, 345-370.

- Borghesi, R., (2008). Widespread Corruption in Sports Gambling: Fact or Fiction? *Southern Economic Journal*, 74, 1063-1069.

- Boyle, G., (2008). Do Financial Incentives Affect the Quality of Expert Performance? Evidence from the Racetrack. *Journal of Gambling Business and*

Economics, 2, 43-60.

- Boyle, G., Guthrie, G., Gorton, L., (2007). Holding Onto Your Horses: Resolving Conflicts of Interest in Asset Management. *Journal of Law and Economics*, forthcoming.

- Fox, K., (2005). *Racing Tribe: Portrait of a British Subculture.* Transaction Publishers.

- Gao, H., (2010). Optimal Compensation Contracts When Managers Can Hedge. *Journal of Financial Economics*, 97, 218-238.

- Gibbons, R., Murphy, K., J., (1992). Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence. *Journal of Political Economy*, 100, 468-505.

- Johnson, J., Bruce, A., Yu, J., (2010). The Ordinal Efficiency of Betting Markets: An Exploded Logit Approach. *Applied Economics*, 42, 3703-3709.

- Laffont, J., J., Martimont, D., (2001). *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press.

- Levitt, S., D., Syverson, C., (2008). Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate Transactions. *Review of Economics and Statistics*, 90, 599-611.

- Rutherford, R., C., Springer, T., M., Yavas, A., (2005). Conflicts Between Principals and Agents: Evidence From Residential Brokerage. *Journal of Financial Economics*, 76, 627-665.

- Scott, M., B., (1968). *The Racing Game*, Aldine Publishing Company.

- Smith, A., (1976). The Wealth of Nations. The Modern Library, New York.

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

Chapter 5

A Note on Market Efficiency Without Noise-Trader Risk

Abstract

It is said that arbitrage is central to the maintenance of an efficient market, where prices reflect the fundamental value of an asset. When the prices of two equivalent assets diverge, an arbitrageur will bet on a convergence and collect a risk-free profit. However, as the limits to arbitrage literature points out, noise traders can drive prices further from fundamentals, and arbitrageurs may therefore ration the resources they commit. In this chapter we examine arbitrage opportunities in a betting market where, crucially, there is no noise-trader risk. We find that two limits to arbitrage remain. First, transaction costs render a significant proportion of arbitrage opportunities unprofitable. Second, collateral requirements lend an element of irreversibility to the enactment of an arbitrage, and therefore mean that a high return must be exacted for the arbitrageur to exercise their 'option to arbitrage'. This is the first empirical evidence that collateral requirements do indeed limit arbitrage. We propose an alteration to the design of this particular market to remove the two limits, and thereby foster market efficiency.

5.1 Introduction

It is said that arbitrage is central to the maintenance of an efficient market, where prices reflect the fundamental value of an asset. When the prices of two equivalent assets diverge, an arbitrageur will bet on a convergence and collect a risk-free profit. As these positive returns come without risk, the arbitrageur experiences no frictions in obtaining capital, and can therefore rectify any market inefficiencies.

In reality, however, there are few risk-free arbitrages. The failure and subsequent bailout of Long Term Capital Management (LTCM), a highly leveraged arbitrage firm, in 1998, demonstrates this vividly. An arbitrageur can lose money, at least in the short run, and given high leverage and collateral constraints, an arbitrage strategy can amount to no more than 'picking up nickels in front of bulldozers.[1]

The limits to arbitrage literature (see Gromb and Vayanos (2010) for a recent survey) highlights a number of impediments to the operation of arbitrage. Most prominent of these is noise-trader, or non-fundamentals, risk. In the model of De Long et al. (1990), irrational noise traders cause prices to deviate from the fundamental value of an asset. An arbitrageur is limited in taking an offsetting position as they have a finite horizon, and the divergence may initially increase, causing them to exit the market at a loss. In other words, the possibility of losing money in the short-run can deter arbitrageurs from participating in the market.

This premise is extended to include agency costs in Shleifer and Vishny (1997). In this model arbitrageurs use investor's capital to trade. As losing money in the short-run can cause investors to remove their funding, arbitrageurs may themselves ration the positions they take, so as not to be left with insufficient funding when a later arbitrage trade can generate higher returns.[2]

In this chapter we analyse arbitrage opportunities in a betting market where there

[1] A description of LTCM trading strategies from a rival money manager (Lowenstein (2001)).

[2] Shleifer and Vishny (1997) also posit that there is an intellectual limit to arbitrage which prevents arbitrageurs from crossing markets to rectify market inefficiencies. Duarte et al. (2006) find evidence of such a limit to arbitrage in fixed income markets. Strategies which require more 'intellectual capital' to implement, display returns above the market benchmark after controlling for risk.

is no noise trader risk. Bets are traded on Betfair, a betting exchange, on the winner of a tennis match. They are also traded on the score in sets. It is therefore possible to replicate the bet on a player to win with a series of bets in the set market. An arbitrageur can, for example, take a long position on a player to win, and take a short position on all the possible scores in the set market. These markets are short-term and do not rely on the convergence typically required for financial market arbitrages. The arbitrageur can lock-in a profit regardless of subsequent moves in the price, and regardless of the outcome of the match.

To give the greatest possibility that anomalies will arise, we examine arbitrage opportunities during matches. We find that arbitrage opportunities are frequent, allow for significant returns given the time frame, and in some cases persist. One possible reason for this is that bettors are unable to continuously monitor markets. Hens *et al* (2006) show that limited market participation means that arbitrage opportunities will be unexploited with probability one. However, a substantial proportion of betting on Betfair is carried out algorithmically. Betfair provide an *Application Programming Interface* to encourage the use of automated betting strategies, which should make it cost-effective to monitor markets continuously and quickly exploit arbitrage opportunities.

We find that two legitimate limits to arbitrage remain. Firstly, transaction costs play a part. Bettors must pay a commission on their net winnings in each market, rendering some arbitrage opportunities unprofitable.[3] Second, collateral must be posted in order to take advantage of any arbitrage. This collateral is not returned until the end of the match, by which time larger arbitrage opportunities may have arisen. Collateral requirements lend an element of irreversibility to the enactment of an arbitrage, and could therefore mean that a high return must be exacted for the arbitrageur to exercise their 'option to arbitrage'. The requirement to post collateral also impedes the arbitrageur in Gromb and Vayanos (2002), Liu and Longstaff (2004) and Kondor (2009);

[3]This effect is found elsewhere. Rosenthal and Young (1990) examine the pricing efficiency of so-called Siamese-twin stocks, which have clearly specified proportional rights to a parent company's cash-flows. It was found that transaction costs limit the returns to an arbitrage of pricing discrepancies between the two stocks.

models in which the arbitrageur can lose money in the short-run. We find persistent arbitrage opportunities without any significant short-term risk. This the first piece of empirical evidence that collateral requirements do indeed limit arbitrage.

In light of the two limits observed we propose two alterations to the design of this particular market. Firstly, at present, transaction costs take the form of commission charged on the net winnings on a market. As the win and set-betting markets are classified as separate markets, a significant proportion of arbitrage opportunities are no longer profitable once commission is included. This would suggest that arbitrage would be encouraged by classifying any markets related to the same sporting event as a single market. In a similar fashion, bettors are required to post collateral for their liabilities in both the win market and the set market, even though liabilities in one may be offset with a position in the other. This would suggest that classifying any markets related to the same sporting event as a single market would also eliminate the collateral requirement limit to arbitrage.

Other authors have looked at arbitrage in betting markets. In a recent paper, Marshall (2009) examined arbitrage opportunities in internet sport betting with bookmakers. He found that arbitrage opportunities offered median returns of 1.5%, and lasted for a median duration of 15 minutes.[4] The novelty of our data is that our arbitrageurs can operate on one exchange, rather than requiring accounts with a series of bookmakers. In addition, the option of taking a short position on a bet, offered by betting exchanges, means that arbitrage opportunities arise more frequently. This is particularly apparent during matches, when meaningful information is constantly arriving and affecting prices.

Finally, bookmakers have the legal right in many jurisdictions to cancel any bet at any time. If one element of an arbitrage trade is cancelled, the arbitrageur may be forced to replace the cancelled bet with another bet which renders the arbitrage

[4]For other examples of arbitrage in betting markets, see Hausch and Ziemba (1990), Edelman and O' Brian (2004), and Lane and Ziemba (2004). The efficiency of tennis betting markets, with respect to expected returns, is considered in Forrest and McHale (2007). For a general survey of betting market efficiency, see Vaughan Williams (2005).

unprofitable. On the betting exchange used here, those providing liquidity have only a 1-5 second window to cancel a bet.[5] This means that any effect of cancellation risk on the enaction of arbitrage should be negligible.

This chapter is structured as follows. In section 6.2 we describe an illustrative arbitrage strategy. Section 6.3 outlines the data and includes summary statistics of the arbitrage opportunities. Section 6.4 concludes. All proofs of arbitrage strategies are in the Appendix.

5.2 Example: Arbitrage Strategy

Bets on Betfair, a prominent U.K. betting exchange, can be traded on a player to win a tennis match and also the score by which the player wins. Bets take the form of a back bet (where the odds plus stake are returned if the outcome occurs) and a lay bet (where the bettor is liable for the odds if the outcome occurs, but pockets the stake if it does not).

Our example strategy involves backing a player to win, and simultaneously laying all the possible set-betting outcomes involved in that player winning. Consider a best-of-five set match. The player can win 3-0, 3-1, 3-2, or lose. Suppose that the back odds on a win are B_W, and the lay odds on a 3-0 win, 3-1 win, and 3-2 win are L_0, L_1 and L_2 respectively.

Suppose also, for initial simplicity, that commission is paid on the net winnings from this arbitrage. In order to guarantee an equal return, whatever the outcome of the match, an arbitrageur should lay $\frac{B_W+1}{L_0+1}$, $\frac{B_W+1}{L_1+1}$ and $\frac{B_W+1}{L_2+1}$ units on the player to win 3-0, 3-1 and 3-2 respectively, for every 1 unit placed on the player to win.

In fact, commission is paid on net winnings for each market where the win and set markets are classified separately. The initial commission charged to Betfair customers is 5%, but this can decline to as low as 2% dependent on the frequency of the customer's betting. We denote the commission rate as $c \in [0.02, 0.05]$.

[5] http://help.betfair.com/contents/itemId/i65767339/index.en.html

In order to guarantee an identical return no matter what the outcome of the match, the arbitrageur should lay $x = \frac{a+(f+zg+z)b}{1-gb}$, $y = \frac{f+(a+zb+z)g}{1-gb}$ and $z = \frac{h+\frac{(a+fb+f+ag)j}{1-gb}}{1-\frac{((g+1)b-(b+1)g)j}{1-gb}}$ on the player to win 3-0, 3-1 and 3-2 respectively - where $a = \frac{B_W(1-c)+1}{L_0+1-c}$, $b = \frac{c}{L_0+1-c}$, $f = \frac{B_W(1-c)+1}{L_1+1-c}$, $g = \frac{c}{L_1+1-c}$, $h = \frac{B_W(1-c)+1}{L_2+1-c}$ and $j = \frac{c}{L_2+1-c}$ - for every 1 unit on the player to win. The returns for this arbitrage strategy are $r = \frac{(x+y+z)(1-c)-w}{w+xL_0+yL_1+zL_2}$.

In what follows, we consider the returns to this arbitrage and also the opposite trade, where the set outcomes are backed and the win outcome is laid. The proofs for both arbitrage strategies are in the Appendix.

5.3 Data

Betting pricing data was obtained from Fracsoft, a company contracted to market Betfair historical pricing data. This data includes the second-by-second quoted odds, and associated volumes, for betting in a range of tennis matches. For our study we looked at data on 13 matches from the quarter-final stage onwards of the Men's Wimbledon Tennis Championships in 2008 and 2009.[6] These matches were chosen as they were likely to display a high degree of liquidity. There are four arbitrage strategies per match so we evaluated 52 arbitrage strategies over a cumulative total of more than 77 hours of play.

To give an example of the analysis of arbitrage opportunities, consider the Final in 2008, which was contested by Roger Federer (ranked number 1 in the world) and Rafael Nadal (ranked number 2). As an indication of the betting interest in this match, Betfair matched GBP 28,334,894 on Nadal to win and GBP 20,802,434 on Federer to win by the end of the match.

For this match we looked at the returns to four arbitrage strategies. 1) back Nadal in the win market, lay Nadal in the set market; 2) back Nadal in the set market, lay Nadal in the win market; 3) back Federer in the win market, lay Federer in the set market; and 4) back Federer in the set market, lay Federer in the win market.

[6]Data on a fourth quarter-final in 2008 was unavailable.

We examined the returns to arbitrage whilst all three possible set outcomes were still possible. As liquidity builds up prior to a match, and steadily declines as the number of possible outcomes falls, this is the most liquid time of the market, and gives us the greatest possibility of arbitrage opportunities. For the first two strategies, this window lasts 2972 seconds. For the last two strategies, this window lasts 14328 seconds. The reason for this difference is that Nadal won the first set after 2972 seconds, thereby ruling out a 3-0 Federer win, whilst Federer did not win a set until the third set, after 14328 seconds, thereby ruling out a 3-0 Nadal win.

Table 6.1 details the summary statistics of arbitrage opportunities for the four strategies outlined above. Frequency of arbitrage is calculated as the ratio of the number of seconds where there is an arbitrage opportunity, divided by the number of seconds where there is liquidity (i.e. at least one quote) for all of the 4 betting options involved in the arbitrage. The arbitrage returns are described for commission levels of 0%, 2% and 5%. A 5% commission on net winnings is charged for those customers who have not accrued any loyalty points with Betfair. It is possible for the most frequent bettors to reduce the commission level they pay to 2%.

0%								
Strategy	Freq. (%)	Returns (%)			Duration (seconds)			N
		Median	IQ Range	Max	Median	IQ Range	Max	
1	36	0.47	0.9	4.9	7	14	575	14236
2	4.1	1.1	1.6	10	7.5	13	110	14236
3	23	0.42	0.66	3.9	8	19	120	2970
4	22	0.92	1.4	5.2	10	16	239	2970
2%								
Strategy	Freq. (%)	Returns (%)			Duration (seconds)			N
		Median	IQ Range	Max	Median	IQ Range	Max	
1	22	0.58	0.87	4.6	6	11	307	14236
2	2.1	0.98	1.9	8.9	4	6	47	14236
3	14	0.4	0.7	3.5	10	14	75	2970
4	10	0.84	1.2	4.1	7	12	135	2970
5%								
Strategy	Freq. (%)	Returns (%)			Duration (seconds)			N
		Median	IQ Range	Max	Median	IQ Range	Max	
1	13	0.58	0.74	4.2	5	10	92	14236
2	0.81	1	1.7	7.1	5	8.5	33	14236
3	5.9	0.39	0.37	3	4	6	63	2970
4	2.2	0.31	0.25	2.4	9.5	6.8	17	2970

Table 6.1. The summary statistics of arbitrage opportunities for strategies 1 to 4 during the Nadal Federer Wimbledon Final in 2008, for commission levels of 0%, 2%, and 5%. Data is taken while all three set outcomes are still possible. Frequency is calculated as the ratio of the number of seconds where there is an arbitrage opportunity, divided by the number of seconds where there is liquidity (i.e. at least one quote) for all of the 4 betting options involved in the arbitrage.

We find that a 5% commission level renders a substantial proportion of arbitrage opportunities unprofitable. For example, in strategy 4, arbitrage opportunities occur with a frequency of 22% if commission is 0%, and with a frequency of 10% and 2.2% for commissions of 2% and 5% respectively. The charging of commission substantially decreases the pool of possible arbitrages, but opportunities are still relatively frequent, especially for regular users of Betfair.

In Table 6.2 we describe the duration and returns to each arbitrage strategy for the full 13 matches, assuming a commission level of 2%, as arbitrageurs are likely to be regular users of Betfair. Arbitrage strategies arise in all 13 matches studied, and 39 of the 52 strategies produce an arbitrage opportunity. Of these 39 strategies, median returns to arbitrage range from 0.0039% (Lopez/Safin strategy 3) to 1.205% (Djokovic/Hass strategy 1). Given the short-time frame, and the frequency with which new opportunities arise, these returns are economically significant. Certain arbitrage opportunities yield returns of 8.5% (Murray/Nadal strategy 3) and 8.9% (Nadal/Federer strategy 2).

The arbitrage opportunities described have a median duration ranging from 1 second (Roddick/Hewitt strategy 3) to 62 seconds (Lopez/Safin strategy 1). Certain arbitrage strategies last up to 163 seconds (Lopez/Safin strategy 2) and 307 seconds (Nadal/Federer strategy 1). The execution speed for an algorithmic trader, using the *Application Programming Interface* provided by Betfair, will vary dependent on their location and the speed of their computer. The consensus on the Betfair Developers Program Forum (http://forum.bdp.betfair.com/) and related blogs is that a U.K. resident should be able to execute trades in 125-500 milliseconds, which is comfortably below the median duration of arbitrage opportunities observed in our study.

Match	Strategy	Freq. (%)	Returns (%)			Duration (seconds)			N
			Median	IQ Range	Max	Median	IQ Range	Max	
2008									
Nadal/Federer	1	22	0.58	0.87	4.6	6	11	307	14236
	2	2.1	0.98	1.9	8.9	4	6	47	14236
	3	14	0.4	0.7	3.5	10	14	75	2970
	4	10	0.84	1.2	4.1	7	12	135	2970
Nadal/Schuettler	1	13	0.48	0.7	6.5	7	8	35	7470
	2	3.4	1.3	1.6	11	5.5	9	34	7470
	3	0	0	0	0	0	0	0	674
	4	0	0	0	0	0	0	0	674
Federer/Safin	1	8.2	0.24	0.42	1.3	6	15	85	6187
	2	0.39	1.2	2.2	2.6	12	3	15	6187
	3	0	0	0	0	0	0	0	1515
	4	0	0	0	0	0	0	0	1515
Murray/Nadal	1	1.6	0.23	0.13	0.28	10	5.5	12	2149
	2	0	0	0	0	0	0	0	2149
	3	19	0.4	0.55	8.5	9	12	109	6968
	4	0.06	0.75	0.58	1	4	0	4	6968
Ancic/Federer	1	0	0	0	0	0	0	0	1224
	2	0	0	0	0	0	0	0	1224
	3	2.4	0.39	0.58	2.1	20	23	110	14159
	4	0.01	0.47	0	0.47	2	0	2	14159
Lopez/Safin	1	4.1	0.52	0.88	4.5	62	63	120	12658
	2	2.2	0.51	1.4	2.4	18	11	163	12658
	3	0.77	0	0.09	0.22	2	24	50	10254
	4	0	0	0	0	0	0	0	10254
2009									
Roddick/Federer	1	6.45	0.19	0.15	0.77	4	6	67	5032
	2	0.28	0.11	0.73	1.3	4	2	7	5032
	3	2.46	0.12	0.07	0.57	2	3	9	2395
	4	3.97	0.29	0.26	3.08	5	4.75	36	2395
Murray/Roddick	1	2.2	0.02	0.13	0.88	7	9.75	31	2259
	2	5.53	0.63	0.62	2.12	4	13.25	37	2259
	3	1.88	0.22	0.36	0.75	3	4	26	4333
	4	4.62	0.63	0.83	2.6	8	10.75	73	4333
Federer/Hass	1	1.5	0.1	0.13	0.45	4	3.25	13	7370
	2	1.6	0.36	0.37	2.89	7	5.5	22	7370
	3	1.11	0.04	0.04	0.04	7	11.5	14	2805
	4	0	0	0	0	0	0	0	2805
Karlovic/Federer	1	1.35	0.11	0	0.11	7	0	7	1352
	2	0	0	0	0	0	0	0	1352
	3	2.03	0.53	0.36	0.84	16	10.75	27	6186
	4	0.18	0.30	2.33	2.62	3	1	5	6186
Roddick/Hewitt	1	6.42	0.58	0.74	2.58	6	6.5	21	5334
	2	0.09	0.43	0.37	0.43	3	0.50	3	5334
	3	26.09	0.63	0	0.63	1	0	1	1646
	4	0	0	0	0	0	0	0	1646
Djokovic/Hass	1	1.61	1.21	0.5	1.97	12	0	12	2744
	2	0	0	0	0	0	0	0	2744
	3	0.24	0.76	0.31	0.87	3	3	8	7879
	4	3.69	0.64	0.81	4.85	6	10.75	73	7879
Murray/Ferrero	1	3.9	0.22	0.25	1.58	4	8.25	63	6140
	2	1.49	0.74	0.32	3.29	7	6	19	6140
	3	0	0	0	0	0	0	0	2726
	4	16.87	0.61	0.74	2.02	6	48	144	2726

Table 6.2. The summary statistics of arbitrage opportunities - for 6 matches from the 2008 Wimbledon Championships and 7 matches from the 2009 Championships - for an assumed working commission level of 2%. Data is taken while all three set outcomes are still

possible. For each match, strategy 1 (3) is to back the first (second) player in the win market and lay the same player in the set market, and strategy 2 (4) is to lay the first (second) player in the win market and back the same player in the set market. Frequency is calculated as the ratio of the number of seconds where there is an arbitrage opportunity, divided by the number of seconds where there is liquidity (i.e. at least one quote) for all of the 4 betting options involved in the arbitrage.

The puzzle remains, therefore, as to why some opportunities are left, at least temporarily, on the table. The remaining limit to arbitrage is the requirement to post collateral in separate markets. An arbitrageur has only a finite amount of capital to post as collateral, so may defer taking certain arbitrage opportunities because of the possibility of a greater opportunity in the future. As arbitrage opportunities are frequent in this environment, it would be reasonable for an arbitrageur to expect another, possibly better, opportunity to arise before the end of the market.

At present an arbitrageur must pay commission, and post collateral for their liabilities, in both the set and the win market. If all markets related to the same sporting event were classified as a single market, with commission and collateral contingent on the net revenues/liabilities on this sporting event, this would remove the remaining two limits to arbitrage. The arbitrageur could take an opportunity without losing the ability to take any that subsequently arise. It would be interesting to see whether market efficiency is then realised.

5.4 Conclusion

The limits to arbitrage literature describes the structural impediments which prevent arbitrage from maintaining market efficiency. The most prominent limit to arbitrage is noise trader risk; where prices can move further from the fundamental value of an asset, at least in the short-run, discouraging the actions of arbitrageurs.

In this chapter we examined arbitrage opportunities in a market where there is no noise trader risk. According to the literature there are now only two possible limits to arbitrage in this environment. A transaction cost, payable in the form of commission on winnings, limits the profitability of arbitrage. We are able to quantify this transaction cost, and therefore we are left with one possible explanation for the arbitrage opportunities that persist: collateral requirements.

As collateral would be tied up until the end of the market, the arbitrageur may require a larger arbitrage return now in order to forsake the option to use the same collateral in a subsequent arbitrage opportunity. In this sense, the decision to arbitrage is like any other decision to invest under uncertainty (see Dixit and Pindyck (1994)). Deciding to invest when the NPV of an arbitrage is greater than zero ignores the irreversibility that collateral requirements inject into the decision to arbitrage.

References

- 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.
- Dixit, A., K., Pindyck, R., S., (1994). *Investment Under Uncertainty*, Princeton University Press.
- Duarte, J., Longstaff, F., A., Yu, F., (2006). Risk and Return in Fixed Income Arbitrage: Nickels in Front of a Steamroller? *Review of Financial Studies*, 20, 769-811.
- Edelman, D., C., O'Brian, N., R., (2004). Tote Arbitrage and Lock Opportunities in Racetrack Betting. *European Journal of Finance*, 10, 370-378.
- Forrest, D., McHale, I., (2007). Anyone for Tennis (Betting)? *European Journal of Finance*, 13, 751-768.
- Gromb, D., Vayanos, D., (2002). Equilibrium and Welfare in Markets With Financially Constrained Arbitrageurs. *Journal of Financial Economics*, 66, 361-407.
- Gromb, D., Vayanos, D., (2010). Limits of Arbitrage: The State of the Theory. Paul Woolley Centre Working Paper.
- Hausch, D., B., Ziemba, W., T., (1990). Arbitrage Strategies for Cross-Track Betting on Major Horse Races. *Journal of Business*, 63, 61-78.

- Hens, T., Herings, J., J., Predtetchinskii, A., (2006). Limits to Arbitrage when Market Participation is Restricted. *Journal of Mathematical Economics*, 42, 556-564.

- Kondor, P., (2009). Risk and Dynamic Arbitrage: The Price Effects of Convergence Trading. *Journal of Finance*, 64, 631-655.

- Lane, D., Ziemba, W., T., (2004). Jai Alai Arbitrage Strategies. *European Journal of Finance*, 10, 353-369.

- Lowenstein, R., (2001). *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. Fourth Estate.

- Liu, J., Longstaff, F., A., (2004). Losing Money on Arbitrage: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities. *Review of Financial Studies*, 17, 611-641.

- Rosenthal, L., Young, C., (1990). The Seemingly Anomalous Price Behaviour of Royal Dutch/Shell and Unilever N.V./PLC. *Journal of Financial Economics*, 26, 123-141.

- Marshall, B., R., (2009). How Quickly is Temporary Market Inefficiency Removed? *Quarterly Review of Economics and Finance*, 49, 917-930.

- Shleifer, A., Vishny, R., W., (1997). The Limits to Arbitrage. *Journal of Finance*, 52, 35-55.

- Vaughan Williams, L., (2005). *Information Efficiency in Financial and Betting Markets*. Cambridge University Press.

Appendix

Arbitrage Strategies

Recall that the back odds for a win were B_W, and the lay odds were L_0, L_1, L_2 for a 3-0, 3-1, and 3-2 win respectively. Also recall that the commission, $c \in [0.02, 0.05]$, is paid on the net winnings in the win market, and the net winnings in the set market. If the player in question loses, then the revenue after commission is $(x+y+z)(1-c)-w$. No commission is paid in the win market as a loss is incurred. If the player wins, then commission is only paid in the win market, as the set markets can be expected to yield a net loss. The revenue after commission is $wB_W(1-c)+y+z-xL_0$, $wB_W(1-c)+x+z-yL_1$ and $wB_W(1-c)+x+y-zL_2$ for a 3-0, 3-1 and 3-2 win respectively, subject to $y+z-xL_0 < 0$, $x+z-yL_1 < 0$ and $x+y-zL_2 < 0$ respectively. For our arbitrage we require that the revenue after commission is equal whatever the outcome of the match. After rearranging, $w=1$, $x = \frac{B_W(1-c)+(y+z)c}{L_0+1-c}$, $y = \frac{B_W(1-c)+(x+z)c}{L_1+1-c}$ and $z = \frac{B_W(1-c)+(x+y)c}{L_2+1-c}$. To solve the three simultaneous equations, define $a = \frac{B_W(1-c)+1}{L_0+1-c}$, $b = \frac{c}{L_0+1-c}$, $f = \frac{B_W(1-c)+1}{L_1+1-c}$, $g = \frac{c}{L_1+1-c}$, $h = \frac{B_W(1-c)+1}{L_2+1-c}$ and $j = \frac{c}{L_2+1-c}$. This leaves $x = \frac{a+(f+zg+z)b}{1-gb}$, $y = \frac{f+(a+zb+z)g}{1-gb}$ and $z = \frac{h+\frac{(a+fb+f+ag)j}{1-gb}}{1-\frac{((g+1)b-(b+1)g)j}{1-gb}}$. The collateral required to enact this arbitrage is $w+xL_0+yL_1+zL_2$ so the arbitrage return is $r = \frac{(x+y+z)(1-c)-w}{w+xL_0+yL_1+zL_2}$. □

Now suppose that the opposite arbitrage is carried out. The win bet is laid, and the set bets are backed. The lay odds on offer for a win are L_W, and the back odds are B_0, B_1 and B_2, for a 3-0, 3-1 and 3-2 win respectively. w will be laid on the win bet, with x, y and z backed on the 3-0, 3-1 and 3-2 win respectively. If the player in question loses then the net revenue after commission is $w(1-c)-x-y-z$. The net revenue after commission is $(xB_0-y-z)(1-c)-wL_W$, $(yB_1-x-z)(1-c)-wL_W$ and $(zB_2-x-y)(1-c)-wL_W$ for a 3-0, 3-1 and 3-2 win respectively. In order to secure a uniform return no matter what the outcome of the match, $w=1$, $x = \frac{wL_W+w(1-c)-(y+z)c}{B_0(1-c)+1}$, $y = \frac{wL_W+w(1-c)-(x+z)c}{B_1(1-c)+1}$ and $z = \frac{wL_W+w(1-c)-(x+y)c}{B_2(1-c)+1}$. To solve the three simultaneous equations, set $a = \frac{L_W+(1-c)}{B_0(1-c)+1}$, $b = \frac{c}{B_0(1-c)+1}$, $f = \frac{L_W+(1-c)}{B_1(1-c)+1}$, $g = \frac{c}{B_1(1-c)+1}$, $h = \frac{L_W+(1-c)}{B_2(1-c)+1}$

and $j = \frac{c}{B_2(1-c)+1}$. Then $x = \frac{a-b(f-gz+z)}{1-gb}$, $y = \frac{f-g(a-bz+z)}{1-gb}$ and $z = \frac{h-\frac{(a-fb+f-ag)j}{1-gb}}{1+\frac{2bjg-jb-jg}{1-gb}}$. The collateral required to enact this arbitrage is $wL_W + x + y + z$ so the return for this strategy is $\frac{w(1-c)-x-y-z}{wL_W+x+y+z}$. □