Information Processing Constraints and Asset Mispricing

Alasdair Brown*

University of East Anglia

January 14, 2013

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

Abstract

We use a series of natural quasi-experiments - centred on betting exchange data on the Men's Wimbledon Single's Tennis Championships of 2011 and 2012 - to determine whether information processing constraints are partially responsible for asset mispricing. We categorise the arrival of information during each match as a treatment, and hypothesise that the arrival of information means that traders' information processing constraints suddenly become binding. We then examine the effect on the price movements of two assets (one explicitly traded, and one implicitly traded through a replicating portfolio). We find that the arrival of information during each match leads to substantial mispricing between the two equivalent assets, and that part of this mispricing can be attributed to differences in the frequency with which the two prices are updated. In other words, while traders receive all the necessary raw information to update the value of both of the assets that they trade, constraints on their ability to process this information force them to focus on the implications of the new information for just one asset. This, in turn, leads to substantial, but temporary, mispricing in this simple asset market.

JEL Classification: G02, G12, G14

EFM Codes: 320, 350

Keywords: natural quasi-experiment, limited attention, betting, information processing constraints

1 Introduction

Traders are bombarded with information on the macroeconomy, industrial sectors and on individual firms. This information comes in a variety of forms: newspaper articles, blogs, tweets, meetings, broker phone calls and colleague emails. Even if traders are attentive and receive all of this raw information, it is inevitable that they will be unable to process it effectively, to deduce the implications for all of the assets (including potential assets) in their portfolio. In this paper we investigate whether this *information processing constraint* has an effect on the level of mispricing in asset markets.

We use a series of natural quasi-experiments centred on Betfair betting exchange data from the Men's Wimbledon Single's Tennis Championships of 2011 and 2012. Trading is conveniently divided between pre-match periods (when little or no information arrives), and so-called 'inplay' periods during the match (when information is arriving constantly). We hypothesise that the arrival of information means that traders' information processing constraints suddenly become binding. Whilst before the match, bettors had time to assess and price the likelihood of, for example, a Roger Federer win and a 3-1 Roger Federer win, during the match they are unable to process the implications of new information for both these bets. Faced with this constraint, bettors may choose to update the value of the bet on Roger Federer to win rather than the bet on the specific score by which Federer will win. In other words, despite having all the necessary raw information to price both bets, information processing constraints may mean that the value of certain assets are not updated in a timely fashion, and therefore mispricing is temporarily observed.

To verify this idea, we examine the evolution of the implied win probability of each player in two markets: 1) the win market, where the bet is explicitly priced, and 2) the set market, where the bet on the player to win is implicitly priced (via a replicating portfolio). We find three pieces of evidence consistent with the notion that information processing constraints are a cause of asset mispricing. Firstly, the mispricing (calculated as the absolute difference between the implied win probability in the win and set markets) is substantially higher inplay than pre-match. This is where we believe that the constraint is more likely to be binding. Secondly, we use a difference-in-difference approach to assess the relative frequency of price changes in the two markets during each match. We find that the price in the set market changes much less frequently during play (even after controlling for intransient differences between the two markets and common effects of information arrival). This suggests that a proportion of the mispricing can be attributed to the price not being as regularly updated in the set market. Finally, we verify that the frequent changes in price in the win market are not simply noise, by calculating the price discovery contribution of each market using a variant of the Hasbrouck (1995) methodology. Consistent with the win market becoming the preferred choice of traders when the information processing constraint becomes binding, we find that the win market contributes at least 82% of price discovery during each match.

The implications of limited attention and limited processing power in asset markets has been modelled by, among others, Hirshliefer and Teoh (2003), Peng and Xiong (2006), Huang and Liu (2007) and Mondria (2010).[1] By identifying circumstances where traders have all the necessary information but limited processing power, our results build on the empirical work of Cohen and Lou (2012). Cohen and Lou hypothesised that conglomerate firms - with operations in a number of industrial segments - were 'complicated' in comparison to single industry firms. This is because the proportion of income generated by each industrial segment evolves over time and a forecast of this proportion requires analysis. The implication of this proposition is that industry-specific information would be more rapidly impounded into the price of single industry firms, and the returns of a portfolio of these firms would therefore predict the subsequent returns of conglomerate firms. This was indeed the result that they found. In other words, while investors had all the necessary raw information to accurately price both simple and complicated firms, constraints on their ability to process this information led to temporary mispricings (in their case, over the course of months).

The idea that assets which require more complicated information processing may be mispriced for longer strikes a chord with our results. A bet on a player to win by a certain score is a more precise, and perhaps more complicated, prediction than simply wagering on the same player to win the match. Snowberg and Wolfers (2010) presented betting market evidence that individuals struggle to correctly predict the likelihood of small probability events. In addition, in terms of interpreting new information received during each match, simple rules of thumb may not suffice in the set betting market. If a player wins an individual point, it is reasonable to suggest that the probability that they win the match either stays roughly the same or goes up. Winning this same point, however, may increase the likelihood of winning by a certain set score, but decrease the likelihood of the other two possible set scores. Without a simple rule of thumb, deciphering the impact of each point for each of the set bets may be more difficult than conducting a similar exercise in the win market.

The trading mechanism on the betting exchange studied in this paper resembles a stan-

[1] There is a large recent empirical literature linking investor (in)attention to asset prices (e.g. Barber and Odean (2008), Corwin and Coughenour (2008), DellaVigna and Pollet (2009), Hirshleifer *et al.* (2009, 2011), Louis and Sun (2010), Da *et al.* (2011) and Chakrabarty and Moulton (2012)).

dard limit order book on a financial exchange. Traders can post liquidity (via limit orders) or consume liquidity (via market orders). The payoffs in these assets resemble those of short-maturity zero-recovery fixed-income assets. Bettors receive a fixed amount if they are correct in their predictions, and lose their stake if they are incorrect, much as you would by investing in a bond. This is also a competitive and dynamic market, meaning that, much as in financial markets, there are significant costs to repeatedly mispricing assets. Bettors who make errors in their estimation of a player's win probability can expect to be picked off by other bettors. These three points give our analysis external validity.

There are also, it should be pointed out, advantages in using the exchange for this type of study. Firstly, there is a clear(er) separation between zero-information periods and information periods than could be expected in any financial market. This allows us to classify the arrival of information as a treatment in our natural quasi-experiment, a treatment which pushes information processing ability toward its limit. There are spikes of information arrival in financial markets, but the distinction between information and zero-information periods is less stark than in sports betting. Furthermore, the speed with which events unfold in these matches (over the course of hours) allows us to take a microscopic look at price changes in the presence of information. This is in contrast to the longer-horizon asset-pricing study of Cohen and Lou (2012).

The attraction of high-frequency betting data has also encouraged the recent work of Croxson and Reade (2011) and Choi and Hui (2012), who both examine Betfair pricing during association football (soccer) matches. Croxson and Reade demonstrate that betting exchange prices incorporate the effect of goals in a timely manner. Choi and Hui, in a similar study, categorise the first goal of each match as either surprising (if scored by the underdog team) or expected (if scored by the favourite team). This approach allows for an examination of the way that prior beliefs shape reactions to information. They find that the betting exchange overreacts to highly surprising events but underreacts to less surprising events. In this paper, we focus instead on the differences in reaction times of two different markets, with the aim of shedding light on the role that information processing constraints play in asset mispricing.

The rest of the paper is organised as follows. In Section 2 we describe the betting exchange and present summary statistics. In Section 3 we conduct the empirical analysis and in Section 4 we discuss alternative explanations for the pattern of results. Section 5 concludes.

2 The Betting Exchange

The data in our study is taken from Betfair, a betting exchange in the U.K.. Betfair provide a limit order book for bets on the full spectrum of sporting events, with a particular emphasis on horse-racing, football and, to a slightly lesser extent, tennis. The exchange operates by matching up bettors who wish to 'back' (bet on) an outcome, with those willing to 'lay' (bet against) the same outcome. Those taking the lay position are assuming the role traditionally taken by bookmakers. The exchange operates in a similar fashion to the standard financial exchanges. Specifically, bettors can either submit a market order, which matches up with an offsetting limit order in the book, or submit a limit order, which sits in the book until an offsetting market order arrives. The exchange generates revenue by charging a commission (between 2% and 5%) on the winner's profits.

We collected data on 14 matches from the quarter-final stages to the final of the Wimbledon Men's Singles Championships in 2011 and 2012 (there are 7 matches in our sample from each tournament). Although this may seem a limited number of matches, high-frequency sampling means that there are just shy of 800,000 observations for some of our later regressions. Wimbledon was chosen as it is the most prestigious of the four grand-slam events in tennis, and the latter stages of the tournament were chosen as these proved the most popular matches for betting. The data was purchased from Fracsoft. For each match we randomly chose one player (to ensure independent observations), and the chosen player is disclosed later in Tables 3, 5, and 6. The data we have is time-stamped and includes quoted 'back' and 'lay' prices (odds) sampled each second, for both pre-inplay periods before each match and inplay periods during each match. The data also includes the last transaction price (odds) and the cumulative volume at each second of trading.

The betting on each match takes place in two markets. The first is the win market where bets are traded on the winner. The second is the set market, which allows for betting on the specific score by which each player wins. As matches in the the grand-slam events are conducted on a best-of-five sets basis, the set market comprises 6 possible outcomes (3-0, 3-1, 3-2 to each player). We chose tennis predominantly because it has an extensive inplay period, but also because it is possible to replicate the bet on a player to win with only 3 bets in the set market.

In all cases we infer the implied probability of an outcome by taking the midpoint of the spread. For example, if the back odds on a player to win are O_B and the lay odds are O_L, then the implied win probability is $(1/(O_B + 1) + 1/(O_L + 1))/2$. To calculate the corresponding implied win probability from the set market, we simply sum the implied probabilities for each of the 3 possible set scores by which he could win. For example, take the odds on Andy Murray to win the Final (against Roger Federer) in 2012. At the start of the match (14:10:18), the back (lay) odds on a Murray win were 2 to 1 (2.05 to 1). In other words those who backed at this price would have received 2 GBP (plus their stake) for each 1 GBP they put down, in the case of a win. A bettor laying this outcome would be liable for 2.05 GBP multiplied by the backer's stake, in the case of a Murray win, and would pocket the backer's stake otherwise. The back (lay) odds were 8.8 (9) on a 3-0 Murray win, 7.6 (7.8) on a 3-1 win, and 7 (7.2) on a 3-2 win. The implied win probability is therefore 0.3306 in the win market and 0.3394 in the set market, reflecting a very small mispricing of Murray's prospects at the start of the match.

In the top two panels of Table 1 we describe the summary statistics on implied win probability for the full 14 matches. The average implied win probability in the win market (.690867) is close to the average in the set market (.682561). The relationship appears to be closer pre-inplay, with respective averages of .6928574 and .6933364 (the averages above 0.5 reflect the fact that the favourite was randomly selected in the majority of our 14 matches). In Figure 1 we plot the implied win probability - of Andy Murray in the 2012 Final - as inferred from the win and set markets. Once again, the implied win probabilities track each other closely. One point to make at this stage is that there are a small number of surprising readings of implied win probability in the set market (incidentally, from matches other than the 2012 Final). For example, the maximum reading in the set market is 1.507688. Unlike in the win market, the implied win probability in the set market is not bounded between 0 and 1. While these extreme readings are rare, we do exclude them later in our study to ensure the robustness of our results.

Before we proceed, it is also worth explaining our choice of quoted prices, rather than transaction prices, to calculate mispricings. We use the midpoint of the quoted spread on the exchange for two reasons. Using transaction prices is problematic because the volume in the set market is often lower than that in the win market. As a result, we could be taking a substantially lagged valuation from the set market, and wrongly inferring that there are large mispricings between the two markets. By using quoted prices, we are (as much as possible) comparing contemporaneous valuations. The second reason for not using transaction prices is the presence of 'bid-ask bounce'. Specifically, there could be a situation where the quoted prices did not change but because a back order was swiftly followed by a lay order - and these prices are separated by the spread - we could wrongly infer from transaction prices that valuations have changed substantially. By taking the midpoint of the spread we are not affected by the bounce caused by opposing order flow.

One other reason to use quoted prices is the famous 'favourite-longshot bias'. This is where, on average, returns from bets on favourites exceed those of bets on longshots. This bias was found as far back as Griffith (1949), and has also been observed, albeit to a lesser degree, on Betfair by Smith *et al.* (2006). The foremost explanation for the bias in markets where bettors have a counterparty (including bookmaker markets and betting exchanges, but excluding pari-mutuel markets) is adverse selection (see Shin (1991, 1992, 1993)). Put simply, the cost to a market-maker of losing out to an insider (with advance knowledge of the outcome of the race) is greater when the winner is a longshot (as they must pay out more). The response of the market-maker is to depress the odds on the longshot further below their empirical probability than they would do for the favourite's odds. This would be a concern in our setting if we used transaction prices - which are likely biased toward 'backer-initiated' bets - as the set market implied win probability would regularly exceed the corresponding measure in the win market. However, using the midpoint of the quoted spread likely mitigates this problem, as this measure should approximate the liquidity provider's prior unbiased estimate of the probability of an outcome. Likewise, if the bias is generated by risk-loving bettors (as in Ali (1979)) or over (under) estimation of small (large) probabilities (as in Snowberg and Wolfers (2010)), taking the midpoint of the spread should largely offset these effects. The fact that the average implied win probability is in fact (marginally) higher in the win market than the set market (see 'All' data in Table 1) confirms this impression.

3 Data Analysis

3.1 Mispricing

Our first task is to establish whether the arrival of public information - in our case, the screening of a series of tennis matches - increases the level of mispricing between the win and set markets. To do this, we measure mispricing as the absolute difference between the implied win probability in the win market and the corresponding implied win probability in the set market. The bottom panel of Table 1 describes summary statistics on our measure. We find that the average mispricing between these two markets is quite small (0.02), partly because there are cases where the measure is nil. This is reassuring because, despite the differences in the way implied win probability is calculated (from 1 or 3 set(s) of prices), this demonstrates that the two markets can, at times, come to a consensus on the probability of a player's win.

Staying with the bottom panel of Table 1, we see the first evidence that mispricing is indeed higher when public information is arriving. There is a more than 10 fold increase in average mispricing during inplay periods compared to the same matches pre-inplay. Figure 1 captures this fact vividly. We plot mispricing regarding Andy Murray's implied win probability in the Final of 2012, for both pre-inplay and inplay (beginning at T=17723). Mispricing is visibly higher during the match.

In Table 2 we test this proposition more formally. In all of the regressions in this table we include random effects for the 14 matches in our study, to ensure that any observed effect is widespread. In the first regression we regress mispricing on an indicator variable equalling 1 if the time period is during a match, and 0 otherwise. We find that mispricing is higher during matches, with significance at the 0.1% level. This is a necessary but not sufficient condition for the notion that binding processing constraints are a driver of asset mispricing. The first thing to note, however, is that any measure constructed from quoted prices will undoubtedly be persistent. Therefore in regression 2 we exclude all observations where neither the implied win probability in the win market nor the implied win probability in the set market changed in the last second. This should reduce the extent to which serial correlation in mispricing is driving our result. Even after excluding these observations, we find that mispricing is significantly higher (at the 0.1% level) during matches.[2] We also wish to ensure that mispricing is not caused by the order book recovering from trade in the last period. If it were, mispricing would be a reflection of temporary illiquidity rather than differences in valuation. To deal with this issue, in regression 3 we exclude all observations where there was an order (in either the win or set market) in the preceding second. We find the aforementioned result is robust to this choice of sub-sample, and, in fact, the difference between the coefficients in regressions 1 and 3 suggests that trade during the match reduces rather than increases mispricing (the inplay effect is higher in regression 3). Finally, we mentioned earlier that there were a few readings of implied win probability in the set market that exceeded 1. In the fourth regression of Table 2, we exclude these readings, thereby omitting 7,520 of the 392,944 observations. Once again, we find that mispricing is higher in inplay periods, with significance at the 0.1% level.

The effect of information on mispricing - a more than 10 fold increase during matches judging by the coefficients in the first regression of Table 2 - is particularly striking when we consider that there are likely to be two factors acting as a restraint on mispricing. First of all, if mispricing increases beyond a certain level, arbitrage opportunities arise. An arbitrageur can construct a simple strategy of betting on a player to win in the win market, and betting against the same player in the set market (or vice versa). Mispricing may need to be greater than 5% for the arbitrage trade to be profitable - as Betfair commission is charged separately in the win and set market - but the presence of arbitrageurs should nevertheless act as a restraint on large mispricing between the two markets. The second factor is that the progression of the match resolves uncertainty about the likely winner. At the end of the match, all uncertainty is resolved and mispricing must converge to zero. There is, after all, no more information to process.

[2] The result is also robust to including lagged mispricing as an additional explanatory variable, or using Newey-West standard errors, to control for persistence in the mispricing measure. The results of these two regressions are not tabulated.

One way to assess the extent to which information processing constraints are responsible for this mispricing is to calculate mispricing when there is only one possible set score by which the player could win. If the player sampled either lost the match, or won by 3 sets to 2, there will be periods in which processing the implications of new information for the set bets is equivalent to processing the same information for the win bets. Essentially, in this situation the player can only win 3-2, so set bet pricing requires no additional effort to that already undertaken for the win bets. In regression 5 of Table 2 we add an indicator variable, equalling 1 if the player can only win by one possible set score at that time, and 0 otherwise. This comprises 9.13% of inplay time for the full 14 match sample. In line with our thinking, the level of mispricing is - judging by the size of the coefficients - more than 50% lower during these periods. In other words, when no additional information processing effort is required to price a set bet, the level of mispricing between win and set bets is much lower.

Up until this point we have included random effects for each match to ensure that our results are not driven by particularly high inplay readings of mispricing in a few matches. We would like, however, to ascertain the breadth of the effect that information arrival has on asset mispricing. We do this by running the first regression of Table 2 individually for each of the 14 matches. Table 3 displays the coefficient associated with the inplay indicator for each of these matches. (We estimated these regressions with White heteroskedasticity-consistent standard errors). The inplay effect is positive and significant in all of the 14 matches, with significance at the 0.1% level. To conclude at this stage, information does appear to induce mispricing on this betting exchange. In the next subsection we investigate, in more detail, whether information processing constraints are partially responsible.

3.2 Information Processing Constraints

To examine whether information processing constraints can explain mispricing, we set up a difference-in-difference model. The aim is to assess the frequency of any changes in valuations during each match. The idea behind the regressions that follow is that we must, firstly, control for intransient differences between the win and set market (in terms of volume, prominence, price mechanisms etc.), and, secondly, we must also control for common effects that the arrival of information has on both markets. Once we have controlled for these two factors, we can then isolate the different impact that information has on the frequency of reaction in each market.

In the top panel of Table 4 we regress an indicator variable equalling 1 if the implied win probability changed in the last second, and 0 otherwise, on three explanatory variables. The first explanatory variable is an indicator for whether the market is the win market (to control for intransient differences between the two markets), the second variable is an indicator for whether the match is inplay (to control for the common effect of information on the frequency of price revision), and the third variable is an interaction between the two aforementioned indicators. The interaction term is crucial as this captures any differences in the frequency of the two markets' responses during information arrival. A logit specification is used and random effects for each match are included. There are three results, all significant at the 0.1% level. Firstly, in the baseline pre-match (no information) period, it appears that the set market is more susceptible to changes in valuations as the coefficient associated with the win market indicator is negative. This is perhaps expected, as trade in the set market allows for finer distinctions between different implied win probabilities as this measure is constructed from 3 prices rather than just 1. Secondly, both markets respond more frequently when information arrives as the coefficient associated with the inplay indicator is positive. This is certainly expected, as traders are more likely to revise their valuations when new information arrives. Thirdly, judging by the coefficient associated with the interaction term, it is the win market that is more likely to respond during matches. This evidence is consistent with the notion that the arrival of information means that constraints on traders' capacity to process information suddenly becomes binding. Rather than interpret the implications of the last point for the 8 bets in the two markets (2 in the win market and 6 in the set market), they are only able to process information related to the win bets. When there is a pause inplay - e.g. when players sit down at the change of ends - perhaps they are then able to update their valuations of all the set betting outcomes. This would explain why updating in the set market is much less frequent inplay.

One concern at this point is that changes in valuations may be a reflection of asymmetries in order flow. We mentioned earlier that volume in the win market was typically higher than volume in the set market. In order to ensure that changes in price are reflections of changes in traders' valuations, rather than a result of the transient impact of orders, we also repeated the first regression in Table 4, but this time excluded all observations where an order had taken place in the market concerned in the previous second. As a result, we can focus solely on changes in valuation that are not induced (directly, at least) by orders. These results are presented in regression 2 of Table 4. The effect remains - the set market responds less frequently than the win market - and indeed is stronger for this choice of sub-sample.

Although we have used random effects to incorporate factors idiosyncratic to each match, in Table 5 we repeat our 2 regressions individually for the 14 matches in the sample. In each case, we display the coefficient associated with the interaction between the win market and inplay indicators, in order to capture the relative frequency of price changes during each match. The results that correspond to the first (second) regression in Table 4 are displayed in the top (bottom) panel of Table 5. (All regressions in Table 5 are estimated using White heteroskedasticity-consistent standard errors). Our earlier results are replicated in all of the 14 matches for the first regression and 12 of the 14 matches in the second regression. In each case where statistical significance (at least at the 5% level) is found, the win market responds more often to information inplay.

3.3 Price Discovery

At this stage we have provided two sets of results that are consistent with the notion that information processing constraints are partially responsible for mispricing in this market. The first piece of evidence was that mispricing was higher inplay (when constraints were more likely to be binding), and the second piece of evidence was that this was partially driven by differences in the frequencies with which prices were updated in the two markets examined. Essentially, we could argue that the set market valuation differed from that of the win market inplay because information processing capacity was focused on the latter market.

One gap in this argument, however, is that we cannot be sure at this stage that the frequent price changes in the win market are not simply noise. If limited processing capacity is truly being focused on the win market bets, then we should see that win market price changes lead set market price changes. In other words, only when there is a lull in play (or indeed only when arbitrageurs arrive on the scene) does the set market catch up with the information processing that has already taken place for the win bets.

To verify this idea, we use a variant of the price discovery models of Garbade and Silber (1983) and Hasbrouck (1995). As our setting has only two prices, rather than n as in Hasbrouck (1995), our model more closely resembles that of Garbade and Silber (1983). We estimate the following two regressions:

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

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

W_t is the implied probability of a player winning in the win market, S_t is the implied probability of the same player winning in the set market, and ϵ_t and u_t are error terms. Both implied win probabilities are defined in Section 2. The idea is straightforward. The greater the coefficient β_1, the greater the contribution of the set market to price discovery. A high β_1 would suggest that a mispricing between the two markets at time $t-1$ is corrected by a subsequent price change in the win market $W_t - W_{t-1}$. This would imply that the set market leads the win market in terms of price discovery, and is the location of the initial information processing. On the flipside, however, the greater the coefficient α_1, the greater the price discovery contribution of the win market. A mispricing at time $t-1$ is corrected by a subsequent price change in the set market $S_t - S_{t-1}$. We expect that $\alpha_1 > \beta_1$, as the win market is the more frequent updater of price (see Section 3.2).[3]

Using these estimated coefficients, we calculated the win market contribution to price discovery, defined as $\frac{\alpha_1}{\beta_1+\alpha_1}$. This measure, along with the coefficients for each of the 14 matches, is displayed in Table 6. (All regressions in Table 6 are estimated using White heteroskedasticity-consistent standard errors). Sampling for this data is carried out at 1 minute intervals to allow sufficient time for information to arrive (this leaves time for approximately one point to be played). In line with our hypothesis - that information processing constraints bind inplay and that limited processing power is focused on the win market - we find that the win market is the major (and sometimes sole) contributor to price discovery. The lowest win market contribution (82%) is found in the first Semi-Final of 2012. In some

[3]For more details on price discovery models, as applied to betting exchange data, see Chapter 2 of Brown (2012).

cases, the contribution of the set market is actually negative, implying that on the occasions that the set market leads the win market, it is more often wrong than right (i.e. win market prices subsequently go in the other direction). This provides quite clear evidence that information processing capacity is being focused on the win market. We checked the robustness of our results to varying the sampling interval (10 seconds, 30 seconds, 2 minutes) and also used Newey-West standard errors to account for serial correlation. The results are qualitatively the same so we do not present them here.

The results in this section do beg one question: why do liquidity providers leave quotes, in the set market, that will likely become out-of-date? If they anticipate that they will be busy processing the implications of new information for win market bets, why do they leave quotes in the set market at all? After all, if these quotes do not incorporate all new information, they are in danger of being adversely selected by traders who have deciphered the correct implications for the set bets. One possible explanation is that the probability of being adversely selected is not very large. Perhaps it requires too much processing power for liquidity takers (those submitting orders) to quickly and consistently identify the liquidity providers' mispriced set quotes. Unless the mispricing is very large - in which case a simple arbitrage strategy can be used (see the discussion in Section 3.1) - identifying which set outcomes have become more/less likely and by how much, and then quickly trading on these conclusions, may not be a simple task.

4 Discussion

In this section we will discuss two alternative explanations for the pattern of results described in this paper. The first alternative explanation is the 'gradual information flow' hypothesis (developed by Hong and Stein (1999) with empirical evidence in Hong *et al.* (2000)). The idea behind this model is that information - particularly private information - diffuses gradually across a population. This creates price momentum as the implications of a piece of new information are reflected in asset prices only after a significant lag. On the face of it, this does not seem an appropriate model for sports betting information. Bettors all view the same match - on television or at the stadium - and therefore can expect to be privy to the same information at the same time. This may not, however, be true. Television pictures of sporting events, including tennis, are broadcast with a lag. This means that those watching the tennis at the venue itself will receive each public signal a few seconds before those watching the same match on television. There are rumours that a number of Betfair customers, so-called 'courtsiders', have been exploiting this opportunity by placing bets with advance knowledge of the outcome of the last point (*Guardian* 29th June 2011). If 'courtsiders' are concentrated in the win market, this could explain why the win market leads the set market in price discovery.

The problem with this explanation for the results is that it is consistent with only two of the three results presented in this paper. It is consistent with greater mispricing of bets inplay (in Section 3.1), as information on the progress of the match diffuses gradually across the two markets. It is also consistent with the win market leading the set market in Section 3.3 (if 'courtsiders' do indeed concentrate in the win market). It is not, however, consistent with the quasi-experimental results in Section 3.2. We find that the set market price is not simply updated later, but it is also updated less frequently. This is consistent, to our mind, only with the idea that information processing constraints force the periodic neglect of the set market.

A second possible explanation is that different interpretations of public information are driving our results. Harris and Raviv (1993) and Kandel and Pearson (1995) present models where traders with different likelihood functions differ in their interpretation of public information. This explanation may be consistent with the mispricing effect in Section 3.1, as information arrival could induce disagreement between the two markets (if they have slightly different trading populations). It is not consistent, however, with the other two results. There is no reason why different interpretations should lead to differences in the frequencies with which traders respond to information (unless one set of traders is utterly unresponsive to certain information). In addition, different interpretations cannot create the price discovery relationship documented in Section 3.3 unless traders learn from the beliefs of others. This learning is not permitted in the models of Harris and Raviv (1993) and Kandel and Pearson (1995) as otherwise disagreement would immediately disappear.

5 Conclusion

Limited attention has become a feature of models of investor learning behaviour (Peng and Xiong (2006)) and portfolio allocation (Huang and Liu (2007) and Mondria (2010)). Some of the more prominent empirical puzzles and anomalies in financial markets - such as stock price momentum and the underdiversification of investor portfolios - can be generated in these models. Moreover, these propositions rely on an extremely uncontroversial assumption: traders do not have unlimited capacity to receive and process information.

In this paper we examine the evidence that information processing constraints lead to asset mispricing. In the process of our examination, we exploit the unique conditions present on a U.K. sports betting exchange. Assets - contingent on the outcome of a tennis match - are traded in two markets (the win and the set market) both pre-inplay (when no information arrives) and inplay (when information is turned on like a tap). We argue that the arrival of information means that traders' information processing constraints suddenly become binding.

We then present three pieces of evidence consistent with the notion that information processing constraints are a cause of mispricing. i) The level of mispricing is 10 times greater during the arrival of information compared to zero-information periods. This is the time when information processing constraints are more likely to be binding. ii) Part of this mispricing can be attributed to differences in the frequency with which traders update the values of the two assets. iii) Price discovery is led by the market with the most frequent updating of prices (during the treatment). This suggests that traders' limited information processing capacity was predominantly put to work in the pricing of just one asset. Other financial market frictions - discussed in Section 4 - fail to capture at least one of these results.

References

- Ali, M., M., (1979). Some Evidence on the Efficiency of a Speculative Market. *Econometrica*, 47, 387-392.

- Barber, B., M., Odean, T., (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial*

Studies, 21, 785-818.

- Brown, A., (2012). *Essays on the Economics of Betting Markets*, SOAS, University of London, PhD Thesis.
- Chakrabarty, B., Moulton, P., C., (2012). Earnings Announcements and Attention Constraints: The Role of Market Design. *Journal of Accounting and Economics*, 53, 612-634.
- Choi, D., Hui, S., K., (2012). The Role of Surprise: Understanding Over- and Underreactions Using In-play Soccer Betting. HKUST/NYU Stern working paper.
- Cohen, L., Lou, D., (2012). Complicated Firms. *Journal of Financial Economics*, 104, 383-400.
- Corwin, S., A., Coughenour, J., F., (2008). Limited Attention and the Allocation of Effort in Securities Trading. *Journal of Finance*, 63, 3031-3067.
- Croxson, K., Reade, J., J., (2011). Information and Efficiency: Goal Arrival in Soccer Betting. University of Birmingham Discussion paper, 11-01.
- Da, Z., Engelberg, J., Gao, P., (2011). In Search of Attention. *Journal of Finance*, 66, 1461-1499.
- DellaVigna, S., Pollet, J., M., (2009). Investor Inattention and Friday Earnings Announcements. *Journal of Finance*, 64, 709-749.
- 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.
- Griffith, R., M., (1949). Odds Adjustments by American Horse-Racing Bettors. *American Journal of Psychology*, 62, 290-294.
- Harris, M., Raviv, A., (1993). Differences of Opinion Make a Horse Race. *Review of Financial Studies*, 6, 473-506.

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

- Hirshleifer, D., Teoh, S., H., (2003). Limited Attention, Information Disclosure, and Financial Reporting. *Journal of Accounting and Economics*, 36, 337-386.

- Hirshleifer, D., Lim, S., S., Teoh, S., H., (2009). Driven to Distraction: Extraneous Events and Underreaction to Earnings News. *Journal of Finance*, 64, 2289-2325.

- Hirshleifer, D., Lim, S., S., Teoh, S., H., (2011). Limited Investor Attention and Stock Market Misreactions to Accounting Information. *Review of Asset Pricing Studies*, 1, 35-73.

- Hong, H., Stein, J., C., (1999). A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets. *Journal of Finance*, 54, 2143-2184.

- Hong, H., Lim, T., Stein, J., C., (2000). Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies. *Journal of Finance*, 55, 265-295.

- Huang, L., Liu, H., (2007). Rational Inattention and Portfolio Selection. *Journal of Finance*, 62, 1999-2040.

- Kandel, E., Pearson, N., D., (1995). Differential Interpretations of Public Signals and Trade in Speculative Markets. *Journal of Political Economy*, 103, 831-872.

- Louis, H., Sun, A., (2010). Investor Inattention and Market Reaction to Merger Announcements. *Management Science*, 56, 1781-1793.

- Mondria, J., (2010). Portfolio Choice, Attention Allocation, and Price Comovement. *Journal of Economic Theory*, 145, 1837-1864.

- Peng, L., Xiong, W., (2006). Investor Attention, Overconfidence and Category Learning. *Journal of Financial Economics*, 80, 563-602.

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

- Shin, H., S., (1992). Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias. *The Economic Journal*, 102, 426-435.
- Shin, H., S., (1993). Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims. *The Economic Journal*, 103, 1141-1153.
- Smith, M., A., Paton, D., Vaughan Williams, L., (2006). Market Efficiency in Person-to-Person Betting. *Economica*, 73, 673-689.
- Snowberg, E., Wolfers, J., (2010). Explaining the Favorite-Long Shot Bias: Is it Risk Love or Misperceptions? *Journal of Political Economy*, 118, 723-746.

Figures

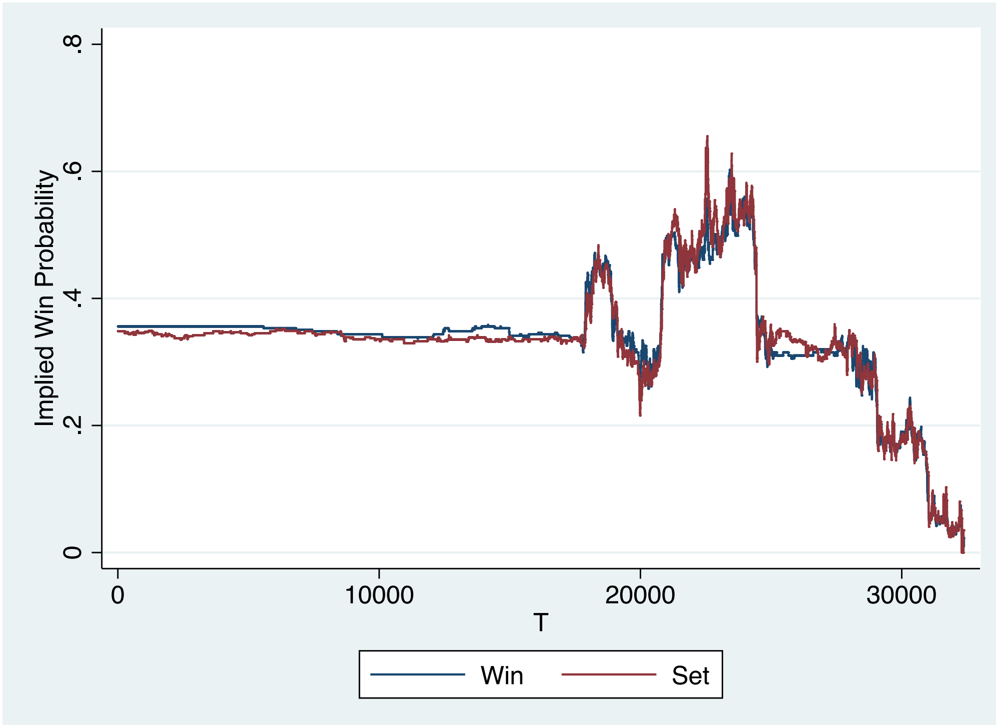

Figure 1: The implied win probability of Andy Murray in the 2012 Wimbledon Men's Singles Championship Final - as inferred from the win market (blue) and the set market (red) - plotted against time. The match began at 14:10:18 (T=17723 on this plot).

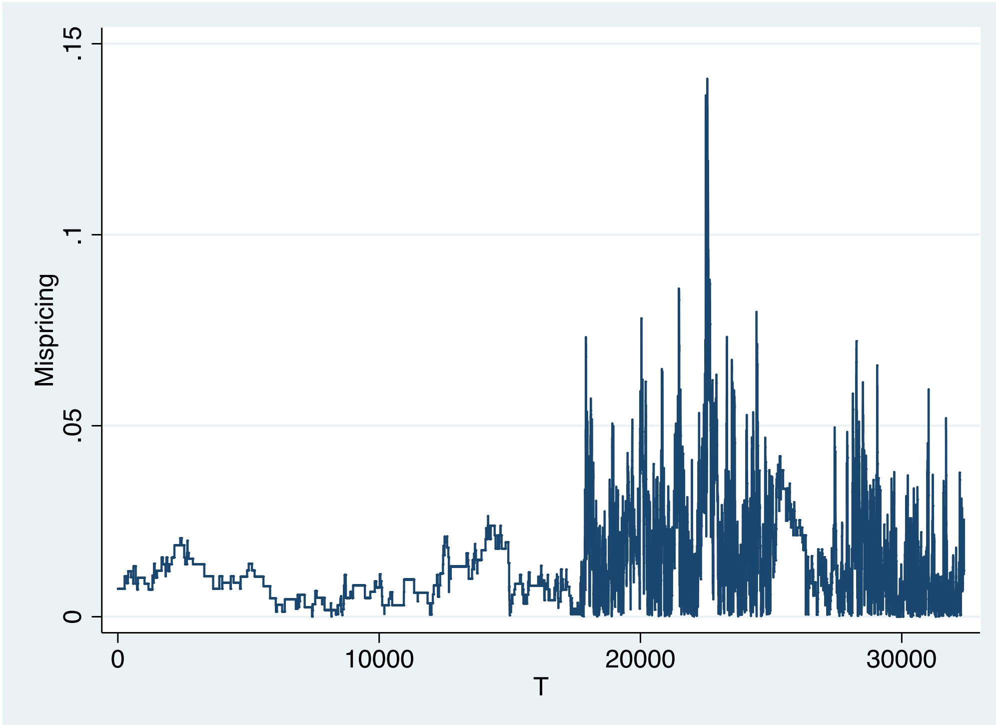

Figure 2: The mispricing of bets on Andy Murray to win the 2012 Wimbledon Men's Singles Championship Final, plotted against time. Mispricing is measured as the absolute difference between the implied win probabilities in the win and set markets. The match began at 14:10:18 (T=17723 on this plot).

Tables

Table 1: Summary Statistics (Implied Win Probability)

Win market	Observations	Mean	Std. Dev.	Min.	Max.
Pre-Inplay	254201	.6928574	.1988313	.2919863	.9569597
Inplay	138786	.6872213	.2807402	.0055	.9852456
All	392987	.690867	.2311134	.0055	.9852456
Set Market	**Observations**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Pre-Inplay	254209	.6933364	.1989548	0	.963456
Inplay	142899	.6633922	.3072107	0	1.507688
All	397108	.682561	.2439415	0	1.507688
Mispricing	**Observations**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Pre-Inplay	254201	.0061245	.0049588	2.32e-06	.3306011
Inplay	138786	.0530399	.1100309	0	.8658171
All	392987	.022693	.0692408	0	.8658171

Summary statistics on implied win probability in the win market, implied win probability in the set market, and the mispricing (absolute difference) between the two. Data is sampled both before and during each match. The data-set comprises the 14 matches from the Quarter-Final stages onwards of the 2011 and 2012 Men's Wimbledon Tennis Championships.

Table 2: Mispricing

Dependent Variable: Mispricing (t)	1	2	3	4	5
	All	$\triangle$ IWP(Win|Set)$\neq 0$	Orders(t-1)=0	IWP(Set)<1	All
Intercept	0.0043192	0.0167905	0.0044452	0.0043159	0.0043245
	(.0062537)	(.0215517)	(.0036837)	(.0063836)	(.0062911)
Inplay Indicator (t)	0.0495418***	0.0429167***	0.0561157***	0.0465738***	0.0522584***
	(.0002059)	(.0014612)	(.0002596)	(.0002073)	(.0002124)
One Set Outcome Indicator (t)					-0.0294461***
					(.0005892)
ρ	0.12901166	0.51238849	0.07111057	0.1368458	0.13107971
No. of Observations	392987	41572	266313	385463	392987
R^2	0.1049	0.0127	0.1424	0.0945	0.1137

A series of regressions to compare mispricing (the absolute difference between implied win probabilities in the win & set markets) in pre-inplay and inplay periods. An indicator variable for inplay is the main explanatory variable. Regression 2 is limited to observations where the implied win probability changed (in either market) in the previous second, Regression 3 excludes observations where an order arrived (in either market) in the previous second, and Regression 4 includes only observations where the implied win probability in the set market is less than 1, to exclude spurious readings of mispricing. In Regression 5 an indicator variable - equalling 1 if there is only one set outcome possible (i.e. 3-2) - is added to the specification. All 5 regressions include random effects for each match, and ρ captures the proportion of variance in the dependent variable captured by these random effects. Standard errors are in parentheses, and * indicates significance at the 0.1% level.**

Table 3: Mispricing (Each Match)

Match	Inplay Indicator (t) coefficient	Standard Error	N	R^2
Murray v Federer 2012 Final	0.0096083***	0.000139	32384	0.1461
Tsonga v Murray 2012 Semi-Final	0.0111361***	0.0001683	32309	0.219
Djokovic v Federer 2012 Semi-Final	0.0194546***	0.0002875	20335	0.241
Federer v Youzhny 2012 Quarter-Final	0.0399362***	0.0008172	19681	0.2622
Djokovic v Mayer 2012 Quarter-Final	0.0611478***	0.0008678	21107	0.3457
Murray v Ferrer 2012 Quarter-Final	0.032999***	0.0002968	37880	0.3208
Tsonga v Kohlschreiber 2012 Quarter-Final	0.3008394***	0.0027017	33102	0.4605
Djokovic v Nadal 2011 Final	0.0102964***	0.0001992	19455	0.1352
Nadal v Murray 2011 Semi-Final	0.0188857***	0.0002306	36794	0.2768
Djokovic v Tsonga 2011 Semi-Final	0.0281305***	0.0004562	24509	0.1573
Murray v Lopez 2011 Quarter-Final	0.0255682***	0.0005666	33394	0.1771
Nadal v Fish 2011 Quarter-Final	0.040167***	0.000582	34287	0.2547
Djokovic v Tomic 2011 Quarter-Final	0.0503126***	0.0004834	22938	0.3924
Federer v Tsonga 2011 Quarter-Final	0.026002***	0.0003781	24812	0.1846

A repeat of Regression 1 of Table 2, this time run individually for each of the 14 matches in the data-set. The only coefficient displayed is that associated with the indicator variable for inplay periods. In each match bets on only one player (the first listed) were sampled. All regressions are estimated with White heteroskedasticity-consistent standard errors (in parentheses), and *, ** and * indicates significance at the 5%, 1% and 0.1% levels respectively.**

Table 4: Information Processing Constraints

Dep. Var.: Change in Implied Win Probability (Indicator)	All	If Orders(t-1)=0
Intercept	-4.520144***	-5.579031***
	(.0675464)	(.0800766)
Win Market Indicator	-1.572785***	-2.175892***
	(.0461928)	(.1063822)
Inplay Indicator (t-1)	2.755914***	3.364475***
	(.0207224)	(.0347103)
Win Market Indicator*Inplay Indicator (t-1)	1.729583***	2.294867***
	(.0473131)	(.1077369)
ρ	0.0175126	0.0220511
No. of Observations where Dep. Var.=1	45145	15914
No. of Observations	794214	650456

Two regressions to assess whether the arrival of information inplay causes information processing constraints to bind, and therefore means that the price in one market is updated less frequently. The dependent variable is an indicator variable equalling 1 if the implied probability changed in the last second. The explanatory variables are an indicator for the win market, an indicator for inplay periods, and an interaction between these two indicators. The interaction term is crucial, as it captures the differences in the frequency with which the two markets respond to information (the treatment). In the second regression we exclude all observations which immediately follow an order in the market concerned. Both regressions incorporate random effects for each match, and ρ captures the proportion of variance in the Dep. Var. that can be attributed to these random effects. Standard errors are in parentheses and *, ** and * indicates significance at the 5%, 1% and 0.1% levels respectively.**

Table 5: Information Processing Constraints (Each Match)

Match	Win Market Indicator*Inplay Indicator (t-1) coefficient	Standard Error	N	Pseudo R^2
	All			
Murray v Federer 2012 Final	1.497584***	0.1273677	64766	0.2091
Tsonga v Murray 2012 Semi-Final	2.94272***	0.1625545	64636	0.2366
Djokovic v Federer 2012 Semi-Final	1.682984***	0.182125	40670	0.2271
Federer v Youzhny 2012 Quarter-Final	3.830158***	0.2578386	43006	0.2872
Djokovic v Mayer 2012 Quarter-Final	4.192396***	0.5858891	44428	0.2338
Murray v Ferrer 2012 Quarter-Final	0.9169568***	0.1149961	75766	0.1953
Tsonga v Kohlschreiber 2012 Quarter-Final	2.026934***	0.1691004	66744	0.2909
Djokovic v Nadal 2011 Final	0.4645506*	0.1888983	38912	0.2159
Nadal v Murray 2011 Semi-Final	0.8167729***	0.1403868	73886	0.2403
Djokovic v Tsonga 2011 Semi-Final	1.697413***	0.2571926	49238	0.1963
Murray v Lopez 2011 Quarter-Final	1.723645***	0.2436261	66878	0.2179
Nadal v Fish 2011 Quarter-Final	3.239931***	0.585237	69754	0.2405
Djokovic v Tomic 2011 Quarter-Final	1.596654***	0.2563132	45906	0.184
Federer v Tsonga 2011 Quarter-Final	3.162542***	0.415702	49624	0.17
	If Orders(t-1)=0			
Match	Win Market Indicator*Inplay Indicator (t-1) coefficient	Standard Error	N	Pseudo R^2
Murray v Federer 2012 Final	2.380325***	0.430861	47570	0.2205
Tsonga v Murray 2012 Semi-Final	2.947642***	0.3534875	53803	0.2342
Djokovic v Federer 2012 Semi-Final	3.114907***	0.7344183	29971	0.247
Federer v Youzhny 2012 Quarter-Final	5.29763***	0.591171	39314	0.382
Djokovic v Mayer 2012 Quarter-Final	3.915712***	0.5929925	40813	0.2713
Murray v Ferrer 2012 Quarter-Final	0.9882271***	0.2492407	59502	0.2297
Tsonga v Kohlschreiber 2012 Quarter-Final	2.266875***	0.290828	60004	0.2741
Djokovic v Nadal 2011 Final	0.039718	0.4784036	25184	0.2658
Nadal v Murray 2011 Semi-Final	0.9270525**	0.3057099	57507	0.2944
Djokovic v Tsonga 2011 Semi-Final	1.874459*	0.7271638	36091	0.2553
Murray v Lopez 2011 Quarter-Final	2.34514**	0.7233241	60173	0.2612
Nadal v Fish 2011 Quarter-Final	3.570462***	1.012624	62924	0.2601
Djokovic v Tomic 2011 Quarter-Final	0.5227171	0.396602	37472	0.2206
Federer v Tsonga 2011 Quarter-Final	2.085963**	0.7219557	40128	0.2131

A repeat of the two Regressions in Table 4, this time estimated for each match in the sample. The only coefficient displayed is that associated with the interaction term between the win market indicator and the inplay indicator. This interaction term captures the relative response of the two markets to the treatment (information). In each match bets on only one player (the first listed) were sampled. All regressions are estimated with White heteroskedasticity-consistent standard errors (in parentheses), and *, **, and * indicates significance at the 5%, 1%, and 0.1% levels respectively.**

Table 6: Price Discovery Contribution

Match	β_1	α_1	Win market contribution	N
Murray v Federer 2012 Final	0.0233286	0.3077421**	0.93	244
Tsonga v Murray 2012 Semi-Final	0.1312763	0.6024121***	0.82	139
Djokovic v Federer 2012 Semi-Final	-0.0156917	1.020275***	1.02	167
Federer v Youzhny 2012 Quarter-Final	0.0171214.	0.7204173**	0.98	83
Djokovic v Mayer 2012 Quarter-Final	0.0075639	0.486052***	0.98	107
Murray v Ferrer 2012 Quarter-Final	-0.0247664	0.510142***	1.05	257
Tsonga v Kohlschreiber 2012 Quarter-Final	0.0024792	0.1484792**	0.98	169
Djokovic v Nadal 2011 Final	-0.2852521*	1.043141***	1.38	149
Nadal v Murray 2011 Semi-Final	0.0502479	1.090315***	0.96	178
Djokovic v Tsonga 2011 Semi-Final	-0.0778784*	0.8453572**	1.10	185
Murray v Lopez 2011 Quarter-Final	0.0271175**	0.5637822**	0.95	122
Nadal v Fish 2011 Quarter-Final	0.0028727	0.743369**	1.00	164
Djokovic v Tomic 2011 Quarter-Final	0.0058966	0.5551947***	0.99	161
Federer v Tsonga 2011 Quarter-Final	-0.0900856*	0.7017461***	1.15	189

A table describing the win market contribution to price discovery - defined as $\frac{\alpha_1}{\beta_1+\alpha_1}$ - for each of the 14 matches in the data-set. β_1 and α_1 are estimates from Equations (1) and (2) respectively. Data is sampled at 1 minute intervals, and in each match bets on only one player (the first listed) were sampled. Price discovery contributions of greater than 1 indicate that on the occasions that the set market leads the win market, it is more often wrong than right (i.e. win market prices subsequently go in the other direction). All equations were estimated with White heteroskedasticity-consistent standard errors and ., *, **, and * indicates significance at the 10%, 5%, 1% and 0.1% level respectively.**