

The Wisdom of Large and Small Crowds: Evidence from Repeated Natural Experiments in Sports Betting*

Alasdair Brown[†]

Fuyu Yang[‡]

University of East Anglia

University of East Anglia

June 11, 2018

*We would like to thank the Editors, two anonymous referees, and an audience at DIW Berlin for very helpful comments. Alasdair Brown acknowledges financial support from the British Academy and the Leverhulme Trust (Ref: SG140097). The research presented in this paper was carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia. The usual disclaimer applies.

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

[‡]School of Economics, University of East Anglia, Norwich NR4 7TJ, U.K.. Email: fuyu.yang@uea.ac.uk

Abstract

Prediction markets have proved excellent tools for forecasting, outperforming experts and polls in many settings. But do larger markets, with wider participation, perform better than smaller markets? In this paper we analyse a series of repeated natural experiments in sports betting. The Queen’s Club Tennis Championships are held every year, but every other year the Championships clash with a major soccer tournament. We find that tennis betting prices become significantly less informative when participation rates are adversely affected by the clashing soccer tournament. This suggests that measures which increase prediction market participation may lead to greater forecast accuracy.

JEL Classification: G02, G12, G14

Keywords: prediction markets, forecasting, sports betting, natural experiment

1 Introduction

Prediction markets – where bets are traded on the outcome of future events – have proved excellent tools for forecasting. Forecasts produced by these markets have historically proved more accurate than polls in predicting political races (Vaughan Williams and Reade, 2016), and have proved more accurate than tipsters (and indeed bookmakers) in predicting sporting outcomes (Spann and Skiera, 2009, Franck *et al.*, 2010). Due to this success, many firms now employ internal prediction markets to improve their decision-making (Cowgill and Zitzewitz, 2015).

But how important is the size of these markets? Do larger markets, with wider participation, perform better than smaller markets? This is an important question, because a prediction market can be designed to attract a large group of traders, or can be designed such that a few key informed individuals are encouraged to take part. In a corporate context, firms could decide to limit prediction market participation to insiders, as is often the case, or open up the market to outside consumers and pundits.

On the one hand, bigger is not necessarily better. The actions of noise traders (De Long *et al.*, 1990) – who trade for non-informational reasons and divert asset prices from efficient

values – may mean that a small market, shorn of such individuals, actually performs better. On other other hand, without noise traders and wide participation, the returns to informed trading will be lower, and therefore information will not be collected and incorporated into prices (Black, 1986, and Bessembinder *et al.*, 2015). It is also possible, of course, that information is not simply in the hands of a small number of informed traders – as assumed in Glosten and Milgrom (1985) – but is instead dispersed across a large number of individuals (Galton, 1907, Surowiecki, 2005). Therefore, only by combining the views of a large crowd can prices be efficient and forecasts accurate.

In this paper we analyse a series of repeated natural experiments in sports betting. The Queen’s Club Tennis Championships are held every year in London, but every other year the Championships clash with a major soccer tournament. This soccer tournament – either the World Cup or the European Championships – attracts a substantial number of bettors, and therefore makes for a smaller crowd for tennis betting. Using a sample of more than 13 million Betfair betting exchange prices from 2008 to 2013, we compare the accuracy of tennis betting prices in odd years when there was no soccer tournament, with the accuracy of prices in even years when there was a soccer tournament. We find that tennis betting prices become significantly less accurate – both pre-match and inplay – when participation rates are adversely affected by the clashing soccer tournament.

Our research setting is ideal to compare the relative forecasting success of large and small prediction markets. One prominent obstacle in estimating a causal relationship between market size and price efficiency (forecast accuracy) is the obvious endogeneity of these two variables. Along the lines of Grossman and Stiglitz (1980), a small inefficient market may become large precisely because traders are attracted by the inefficiency. (Until the inefficiency disappears, of course, and these traders then leave). It is therefore important that we have an exogenous shock to market size that has no direct impact on price efficiency. It is also important that our natural experiment is repeated. There is no reason why tennis matches in an odd or even year should be fundamentally more or less difficult to forecast. The only factor that should affect the efficiency of these prices is the effect that the clashing soccer tournament has on participation. We are therefore not concerned about confounding the effect of the clashing soccer tournament with any steady drifts over time in betting market

price accuracy. This can be an issue with a simple before/after comparison.

In fact, the exogenous scheduling of soccer matches has been used as an instrument in two quite different settings. Metcalfe *et al.* (2011) analysed the effect of the same summer soccer tournaments on U.K. student performance, as these tournaments clash with important examinations and revision-time in even years. In addition, Webbink *et al.* (2016) analysed the effect of media attention on the solving of crime, exploiting the fact that in certain weeks a Dutch crime programme clashes with the screening of live European soccer and therefore has lower viewing figures.¹

There has been quite mixed evidence on the effect of market participation on price efficiency in sports betting markets, when other empirical approaches have been used. For example, Lahvička (2014) and Abinzano *et al.* (2016) find that the favourite-longshot bias is more pronounced in high-profile tournaments, when we might expect betting market participation to be higher.² However, these high-profile tournaments differ from lower-profile events not simply in terms of betting interest but also, one might expect, in terms of the difficulty of the underlying forecasting problem. The difficulty of forecasting Queen's Club matches, on the other hand, should not vary across even and odd years: only the exogenous level of betting market participation is likely to vary (due to the clashing soccer tournament). Adding to the mixed results, Flepp *et al.* (2014) find that increased liquidity is associated with weaker price efficiency for weekend soccer matches (when betting market participation should be higher), but not weekday soccer matches. However, in their paper there is quasi-experimental variation in the minimum cost of trading (liquidity), facilitated by Betfair tick size boundaries, but there is no quasi-experimental variation in actual betting market participation. The repeated natural experiments we study in our paper provide precisely that type of quasi-experimental variation in betting market participation.

The rest of the paper is structured as follows. In Section 2 we describe the scheduling of the tennis and soccer tournaments, and describe our data and variables. In Section 3 we

¹Izquierdo Sanchez *et al.* (2016) also use the exogenous scheduling of major soccer tournaments to assess the degree of substitution between leisure activities; they find substantial drops in cinema attendance during these events.

²In addition to these two papers which consider the efficiency of tennis betting markets by the degree of the favourite-longshot bias, others – including Easton and Uylangco (2010) and McHale and Morton (2011) – consider whether calibrated models of tennis match outcomes can produce positive betting returns.

present our analysis. We conclude in Section 4.

2 Data

The Queen’s Club Tennis Championships are one of the oldest male tennis tournaments, dating from 1890. The week-long tournament runs from Monday to Sunday in early June every year, and involves 6 main knockout rounds, preceded by qualifiers. All matches are played in a best of three sets format. The tournament is usually used as grass-court preparation for the Wimbledon Tennis Championships, which follow later in June, and is screened on the British Broadcasting Corporation (BBC) in the U.K.. The BBC screen one or two matches on their main channels, BBC1 and BBC2, and in recent years have screened additional matches online at www.bbc.co.uk. (Incidentally, the availability of third and fourth tier matches online in recent years is one reason why it is important that our natural experiment is repeated and not a simple before and after comparison; the effect of the latter could be confounded with drifts in viewing availability over time).

In even-numbered years, the Queen’s Club Championships clashes with one of two major soccer tournaments: the World Cup or the European Championships. Between 2008 and 2013 (the period of our data), there were European Championships in 2008 and 2012 and a World Cup in 2010. In 2008 and 2012, there were soccer matches on every day of the tennis championships. For example, Queen’s (as it is often referred to) took place from the 9th-15th June in 2008, and the European Championships took place from the 7th to the 29th June. The situation was slightly different in 2010, as the World Cup did not begin until the 11th June, while Queen’s started on the 7th June. Nevertheless, there was a substantial build-up to the soccer tournament – not least involving the last pre-tournament friendly matches – which would have distracted tennis bettors relative to odd years with no soccer tournament.

One issue surrounding non-Grand Slam tournaments, such as Queen’s, is the participation of the top players. The level of betting interest in these tournaments is often dependent on whether the high-ranked players – not least the ‘big-four’ of Roger Federer, Rafael Nadal, Novak Djokovic and Andy Murray – take part. Although the participation of these players is of course not dependent on whether there is a concurrent soccer tournament, their participa-

tion decisions may introduce noise into our natural experiment. Over the period 2008-2013, Roger Federer did not take part at all, Andy Murray competed every year, Rafael Nadal entered in 2008, 2010 and 2011, and Novak Djokovic entered in 2008 and 2010.

To ascertain the significance of this variation in participation, we collected data from tennisdata.co.uk on every Queen's Club match between 2008 and 2013. This data include the identity and ranking of each player at the time of each match. We begin by checking whether the aforementioned 'big-four' participated in more matches in even or odd years. In Panel A of Table 1, we regress an indicator variable equalling 1 if the player was a member of the big-four (and 0 otherwise) on an indicator variable equalling 1 if the match was played in an even year (and 0 if the match was played in an odd year). There are two observations for each match, one for each player. We find that the participation rates of the big-four did not vary significantly across even and odd years. While the big-four may have entered the tournament slightly more frequently in even years, due to early exits they did not participate in significantly more matches.

In Panel B, we repeat the regression but add the year number – from 1 (2008) to 6 (2013) – to our regressions. This controls for any drift in big-four participation over the years. We find that big-four participation declined slightly over the 6 years, and that once we control for this, big-four participation was actually lower in even years (albeit insignificantly so). In the remainder of Table 1 we present similar analysis where we replace the big-four indicator with an indicator equalling 1 if the player was ranked in the top ten at the time, or the ranking of the player. In all cases – whether we control for drifts over time or not – we find no significant differences in the probability of top-ten involvement or the ranking of the players in even or odds years. In short, we find little evidence to suggest that noise in the participation of players across even and odds years affects the internal validity of the natural experiment.

We centre our main analysis on the Betfair betting exchange. Founded in 2000, Betfair is the largest betting exchange in the world, and operates markets across sports, politics and economics. The main function of the exchange is, of course, betting, but these markets operate in very similar ways to predictions markets set up purely for the purpose of forecasting. In fact, recognising the desire of others to use their betting prices as forecasting

tools, Betfair now display market implied probabilities of various events under the banner ‘Betfair Predicts’.³ The exchange operates as a standard limit order book as used in many financial markets (Parlour and Seppi, 2008). This means that traders can post limit orders, which act as quotes for others, or place market orders, which execute against prices quoted by others. Traders can make quotes, but as in bookmaker markets these quotes will not necessarily be matched by another trader. Bettors can take long positions that an outcome (e.g. Andy Murray to win a given match) will occur, with ‘back’ bets, or short positions that an outcome will not occur, with ‘lay’ bets. Prices are quoted in the form of odds plus the stake. For example, Betfair odds of 5 are equivalent to 4-1 in traditional betting terms, which means that 4 GBP will be returned in addition to the stake for every 1 GBP wagered, if the outcome occurs.

We obtained our main data from Fracsoft, a third-party company which until 2014 was licensed to distribute Betfair data. The Fracsoft data provide second-by-second sampling of the Betfair limit order book. This gives the researcher the best three back and lay quotes, with associated volumes, every second, and also the last transaction price and total volume transacted, again measured each second. We obtained data for every year of the Queen’s Club Championships from 2008 to 2013. This data include pre-match betting from the morning of each match until it begins, and inplay betting for the whole duration of each match. In total, we have more than 13 million observations (taken each second) of Betfair prices from 271 matches over the 6 years. 9.6 million of these observations are pre-match, with 3.6 million inplay.

In our upcoming analysis we are interested in three main measures, which we summarise in Table 2. Each of these measures is summarised separately for pre-match betting (Panel A) and inplay betting (Panel B). The first two measures are related to the frequency and size of trade. We define a trade indicator, equalling 1 if there was a trade for that bet in that second and 0 otherwise. This allows us to measure whether trades became less frequent when there was a clashing soccer tournament. Trade size, the second of our measures, is calculated in GBP. Another way of looking at these first two measures is that the trade indicator is a proxy for the extensive margin of betting (i.e. the number of people betting) and trade size

³<https://www.betfairpredicts.com/>.

is a proxy for the intensive margin of betting (i.e. the amount that active participants bet).

Our third measure is the forecast error, which is defined as the absolute difference between the price of the bet (in implied win probability terms using the midpoint of the back-lay spread) and the realized value of the bet, which is 1 if the outcome subsequently occurred and 0 if not. For example, if in one second we observe Betfair midpoint odds of 5 for a player that subsequently goes on to lose, this would translate to a forecast error of 0.2. This is because the implied win probability for Betfair odds of 5 is $1/5 = 0.2$ and the realized value was 0. If the same player had won, the forecast error would have been 0.8. We can see that forecast errors are on average lower during inplay betting. This is to be expected, as the closer we get to the end of each match the closer we get to the realization of the terminal value of the bet.

Before we proceed to our main analysis it is worth outlining, for those not familiar with the game, the ways in which information can be collected for betting on tennis matches. While the types of information that are relevant in financial markets or political prediction markets may be more obvious – e.g. balance sheet data or local polling data respectively – this may not be the case for tennis betting. Prior to matches, a bettor could study the recent form of a player and, in particular, his recent match-ups against his current opponent on the surface they are about to play on (e.g. grass, hard-court, clay). When it comes to inplay betting, one common practice is to go the venues and place bets before those watching at home. These so-called ‘courtsiders’ benefit from knowledge of the outcome of the most recent point, giving them a series of fleeting advantages over those not at the venue. Another way to gain an edge in inplay betting is to devise a model of the outcome of the tennis match based on the current score. The simplest approach would be to assign a fixed probability that player A wins a point on serve (and a different probability on return), and treat each point independently. More complicated models would allow for momentum, fatigue, and the specifics of historical match-ups between the two players. These models are often combined with a court-side position to create the greatest advantage.⁴ In short, there are a number of choices that bettors must make regarding the information they collect, and the effort they put into forecasting the outcome of the tennis match. It is likely that these choices are affected

⁴‘Inside the Shadowy World of High-Speed Tennis Betting’, Carl Bialik, *538*, May 29th 2014.

by the size of these markets and the returns on offer for informed trading, which is likely to be determined, in part, by whether or not there is a concurrent soccer tournament.

In this paper we compare the accuracy of betting prices in even years, when there was a soccer tournament, to the accuracy of prices in odd years, when there was not. But why not also exploit the within-year variation? We might expect that if a soccer match is concurrently being played it would have an even greater impact on the efficiency of tennis prices than if a soccer match was simply scheduled for later in the same day. The problem with exploiting within-year variation is that the tennis scheduler has control over the scheduling of matches within each day. Therefore, he or she may choose to schedule a high-profile match, such as one involving the British player Andy Murray, at a different time from an England soccer match, for example. After all, it would make sense for sponsorship reasons for the Queen's organisers to keep their highest profile matches away from competing entertainment events. But, if this is the case, we could not reasonably conclude that prices were more efficient for the Murray match – relative to a low-profile match on the same day – simply because the England match did not take place at the same time as the Murray match. The Murray match would likely have garnered the attention of a significant number of informed bettors irrespective of the timing of the England match.

Importantly, while the scheduler has discretion within each day – and this leads to the within-year selection effect we just described – there is no such discretion regarding the week in which the tournament is played. Queen's Club Championships is set for the same week every year in the Association of Tennis Professionals (ATP) Calendar, and therefore the clash with soccer tournaments every other year is unavoidable. Therefore, we can reasonably argue that any resultant drop in betting price efficiency in even years is caused by the drop in participation caused by the concurrent soccer tournament.

3 Analysis

We begin our analysis by firstly checking whether the clashing soccer tournament had an impact on the level of participation in tennis betting. This is central to our claim that prediction market crowds were exogenously smaller in years with a soccer tournament. In

Panel A of Table 3, we regress an indicator variable – equalling 1 if there was a trade in that second and 0 otherwise – on an indicator variable equalling 1 if the second sampled was in an even year (with a soccer tournament) and 0 if it was in an odd year (without a soccer tournament). For this and all subsequent regressions we cluster standard errors at the selection level. For example, Andy Murray in the Final in 2011 would be one cluster. We run this regression separately for pre-match and inplay betting. We find that the soccer tournament did indeed have a substantial impact on the frequency of trade: pre-match trade was 18.33% less frequent in even years and inplay trade was 15.15% less frequent in even years. In regressions 3 and 4 in Panel A we then regress the size of trades, measured in GBP, on the even year indicator. Based on our results, there does not seem to be any significant difference in the size of trades across even and odd years, either pre-match or inplay. At this stage, our results suggests that the extensive margin of tennis betting (i.e. the number of participants) declined, but not the intensive margin (i.e. the amount of betting by active participants). Of course, this translates to a decline in overall trading volume.

In Panel B of Table 3 we repeat the analysis in Panel A, but this time add the year number – from 1 (2008) to 6 (2013) – to our regressions. This controls for any drift in the frequency and size of trades over time. In regressions 1-4 of Panel B, we find no evidence of drift, and indeed find qualitatively similar results to those in Panel A. After controlling for any time-trend, the frequency of pre-match trading declined by 18.78% and the frequency of inplay trading declined by 14.73%. There was no significant impact on the size of trades, either pre-match or inplay. Overall, our results in Table 2 suggest that the clashing soccer tournament did indeed cause there to be a smaller crowd for tennis betting.⁵

We next turn to the main part of our analysis: the effect of the clashing soccer tournament on the accuracy of prices. Our first dependent variable is the forecast error, defined as the absolute difference between the price of the bet (in implied win probability terms) and the realized value of the bet, which is 1 if the outcome subsequently occurred and 0 if not. We measure these forecast errors every second, pre-match and inplay. As we can see from Table 2, forecast errors are on average higher pre-match; at the end of each match the forecast

⁵These results are also robust to the addition of fixed effects for day of the week (a proxy for tournament round) and fixed effects for each player.

error will of course be zero. Replicating the structure of Table 3, in Panel A of Table 4 we regress the forecast error on the indicator variable equalling 1 if the price was in an even year (when there was a soccer tournament). We find that forecast errors were significantly higher in even years. In pre-match trading forecast errors were 12.75% higher in even years, and in inplay trading forecast errors were 15.84% higher in even years. In Panel B of Table 4 we again add the year, numbered 1 to 6, as an additional control variable to control for any time-drift in forecast errors. There is only slight evidence of an upward drift in forecast errors in inplay trading, and once we control for this, we find qualitatively similar results to Panel A. Pre-match forecast errors were 15.9% higher in even years, and inplay forecast errors were a startling 23.25% higher in even years. Based on our first measure of forecast accuracy, we find clear evidence that a smaller crowd leads to less efficient prices and less accurate forecasts.

To summarize our forecast error results, in Figure 1 we plot the average forecast error for each year in our sample. This plot runs from 2008 to 2013, and pre-match and inplay forecast errors are displayed separately. With the exception perhaps of 2008, we find a zigzagged pattern that might be expected with repeated natural experiments. Forecast errors go up in years when the soccer tournament is on and the tennis crowd is exogenously smaller, and go back down when there is no soccer tournament and the crowd recovers in size.

There are of course alternative ways to measure the accuracy of betting market prices. One popular method is to measure the degree of the favourite longshot bias. This bias – dating at least as far back as Griffith (1949) – describes the tendency of favourites to win more often than betting prices imply, though in some cases, notably Busche and Hall (1988), the bias is in the opposite direction with longshots winning more often than prices imply. The degree of the bias is measured with the following equation:

$$y_i = \beta_0 + \beta_1 x_{it} + \epsilon_{it} \tag{1}$$

y_i is an indicator variable, equalling 1 if player i won the match, x_{it} is the implied win probability of player i winning as measured from the odds at time t , and ϵ_{it} is an error term. In an ‘efficient’ market, β_1 will have an estimated value close to or equal to 1. This equation is often labelled the Mincer-Zarnowitz regression (Mincer and Zarnowitz, 1969), and has been

used to assess betting price efficiency by Vaughan Williams and Reade (2016) and Brown and Yang (2016), among others.

Before we present our results, it is worth noting that there is some debate about the reasons for the favourite-longshot bias. One interpretation is that deviations of β_1 from 1 reflect mispricing. However, others have linked the bias to risk-loving behaviour, as bettors over-bet longshots, thereby driving down their returns, and under-bet favourites, thereby driving up their returns (Weitzman, 1965). Alternatively, others argue that the bias arises due to a failure to accurately calibrate the probability of low-likelihood events (e.g. Snowberg and Wolfers, 2010), or even due to adverse selection from informed bettors (e.g. Shin, 1993). In short, differences in the deviation of β_1 from 1 in odd and even years may be due to mispricing (which is our focus), or may be due to a range of other factors (see Ottaviani and Sørensen, 2008, for a survey of explanations for the bias).

We estimate Equation (1) and present our results in Table 5. Separate regressions are run for odd and even years, and pre-match and inplay trading. While examining our β_1 estimates, we may also be interested in the R^2 ; the proportion of variation in the win indicator that can be explained by prices (implied win probabilities). The higher the R^2 , the more informative the prices. The R^2 was used to measure betting price efficiency by Franck *et al.* (2010) and Brown and Yang (2016). For our β_1 estimates, we actually find larger deviations from 1 in odd years, suggesting that prices were less efficient when the crowd was exogenously larger. This applies to both pre-match and inplay trading. The R^2 , on the other hand, was higher in odds years – both pre-match and inplay – suggesting that there were more informative prices when the crowd was exogenously larger. In short, the results in Table 4 present quite a mixed picture.

Perhaps the best way to evaluate the informativeness of prices is to use them to forecast the winners and losers of tennis matches, and compare this forecast accuracy across odd and even years. One approach along these lines is to construct Receiver Operating Characteristic (ROC) curves. This is a method originally used to determine the accuracy of radar systems, but subsequently used predominantly in assessing the accuracy of medical diagnostic tests. We begin by estimating logit regressions of a similar form to Equation (1). A win indicator – equalling 1 if the player won the match and 0 otherwise – is regressed on the implied win

probability as inferred from the betting prices. The results of these regressions are displayed in Table 6. We use these estimates to predict winners and losers, and then evaluate these predictions against actual outcomes. When evaluating the accuracy of binary forecasts, we are concerned with the ‘Sensitivity’ of the model – which in our case is the proportion of match winners correctly identified by the model (and, by extension, the betting prices) – and also the ‘Specificity’ of the model, which is the proportion of match losers correctly identified by the model. Greater Sensitivity comes at the cost of lower Specificity as, to take an extreme example, if you said that every player was going to win you would not identify a single loser.

In Figure 2 we plot ROC curves for pre-match and inplay data, in the left and right panel respectively. The method for plotting these curves is described in DeLong *et al.*, (1988). Odd years (without a soccer tournament) and even years (with a soccer tournament) are plotted separately. The curves plot Sensitivity against 1-Specificity. The curve is plotted by selecting a range of cut-offs – e.g. if the model states the player has a win probability >0.55 , then predict a win – and evaluating Sensitivity and Specificity for each cut-off. The area under the curve can be interpreted as the probability that a randomly selected (actual) winner of a match had a higher win probability (according to the model, and therefore the prices) than a randomly selected loser. A model, and the inputs into the model, therefore only has any forecasting accuracy if the area under the ROC curve is greater than 0.5.

We can see from Figure 2 that forecast accuracy was significantly higher in odd years, when market participation was exogenously higher, than in even years. This applies to both pre-match and inplay trading. For pre-match trading, the area under the ROC curve was 0.8033 in even years and 0.8939 in odd years. For inplay trading, the area under the ROC curve was 0.8257 in even years and 0.9093 in odd years. In short, prices were more informative – and helped to produce models that forecasted more accurately – when the prediction market crowd was exogenously larger. Looking at all of our results from Tables 4-6 and Figures 1-2, the overall impression is that larger crowds do indeed make for more accurate prices and forecasts.

Why is that larger prediction market crowds produce more accurate forecasts? As alluded to in the Introduction, there are at least two potential explanations. Large crowds may create higher returns and incentives for informed trading, and therefore lead to more information

being collected, traded upon, and incorporated into prices. In the tennis betting context, this would mean that large crowds incentivize bettors to study the form and match-ups of players, develop new models of inplay bet pricing, and perhaps get a courtside seat in order to gain a fleeting informational advantage over bettors watching at home. The second potential explanation is that information is perhaps dispersed amongst the crowd. Therefore, only by gathering a large crowd can we – by something akin to the law of large numbers – get an unbiased and accurate estimate of a player’s win probability.

Unfortunately, our data do not allow us to convincingly distinguish between these two explanations. We would need to observe the information possessed by each trader, and then ascertain whether the new traders that arrived in the tennis betting market in odd years were informed, or whether they simply increased the liquidity of the market and thereby increased the returns and incentives for informed trading. Our data do not allow us to observe the identity of each bettor, let alone the information they possess and acquire. What we can say is that there is appears to a positive causal relationship between market participation and the accuracy of prediction market prices.

4 Conclusion

Prediction markets – where participants bet on the outcome of future events – have increased in popularity and scope in recent years. The use of such markets is no longer restricted to sports and politics, as firms across a range of sectors now use these markets in an attempt to improve their internal decision-making.

The designer of a prediction market must decide whether to structure the accessibility and the incentives of the market so that as many individuals participate as possible, or limit participation to a few chosen, perhaps more informed, individuals. On the one hand, inviting large numbers of traders may lead to a degree of uninformed ‘noise-trading’, which may distort prices. On the other hand, without the depth and profits created by a large trading crowd, there is less incentive for others to collect and trade upon new information. And, as the popular conception of the ‘wisdom of crowds’ implies, information may actually be dispersed quite widely across the population. Getting more traders in may increase the

accuracy of prediction market prices and forecasts.

In this paper we analyse a series of repeated natural experiments in sports betting to examine the causal effect of market/crowd size on prediction market accuracy. The Queen's Club Tennis Championships are held every year, but every other year the Championships clash with a major soccer tournament. The scheduling of the soccer tournament leads to an exogenously smaller tennis betting crowd. Using more than 13 million Betfair betting prices between 2008 and 2013, we compare the accuracy (or efficiency) of prices in odd years when crowds were relatively large, with the accuracy of prices in even years when crowds were relatively small (due to the soccer tournament). By the majority of our measures, we find that tennis betting prices were significantly more accurate in years with a larger crowd. This applies to both pre-match and inplay trading. This suggests that the designers of prediction markets may wish to open these markets up to a wider crowd in order to improve forecast accuracy.

References

- Abinzano, I., Muga, L., Santamaria, R., (2016). Game, Set and Match: The Favourite-Longshot Bias in Tennis Betting Exchanges. *Applied Economics Letters*, 23, 605-608.
- Bessembinder, H., Hao, J., Zheng, K., (2015). Market Making Contracts, Firm Value, and the IPO Decision. *Journal of Finance*, 70, 1997-2028.
- Black, F., (1986). Noise. *Journal of Finance*, 41, 529-543.
- Brown, A., Yang, F., (2016). Market Liquidity and Price Efficiency: A Field Experiment. Working paper.
- Busche, K., Hall, C., D., (1988). An Exception to the Risk Preference Anomaly. *Journal of Business*, 61, 337-346.
- Cowgill, B., Zitzewitz, E., (2015). Corporate Prediction Markets: Evidence from Google, Ford, and Firm X. *Review of Economic Studies*, 82, 1309-1341.

- 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.
- DeLong, E., R., DeLong, D., M., Clarke-Pearson, D., L., (1988). Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Non-Parametric Approach. *Biometrics*, 44, 837-845.
- Easton, S., Uylangco, K., (2010). Forecasting Outcomes in Tennis Matches Using Within-Match Betting Markets. *International Journal of Forecasting*, 26, 564-575.
- Flepp, R., Nüesch, S., Franck, E., (2014). Liquidity, Market Efficiency and the Influence of Noise Traders: Quasi-Experimental Evidence from the Betting Industry. Working paper.
- Franck, E., Verbeek, E., Nüesch, S., (2010). Prediction Accuracy of Different Market Structures - Bookmakers Versus a Betting Exchange. *International Journal of Forecasting*, 26, 448-459.
- Galton, F., (1907). Vox Populi, *Nature*, 1949, 450-451.
- Glosten, L., R., Milgrom, P., R., (1985). Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics*, 14, 71-100.
- Griffith, R., M., (1949). Odds Adjustments by American Horse-Race Bettors. *The American Journal of Psychology*, 62, 290-294.
- Grossman, S., J., Stiglitz, J., E., (1980). On the Impossibility of Informationally Efficient Markets. *American Economic Review*, 70, 393-408.
- Izquierdo Sanchez, S., Elliott, C., Simmons, R., (2016). Substitution between Leisure Activities: A Quasi-Natural Experiment using Sports Viewing and Cinema Attendance. *Applied Economics*, 48, 3848-3860.
- Lahvička, J., (2014). What Causes the Favourite-Longshot Bias? Further Evidence from Tennis. *Applied Economics Letters*, 21, 90-92.

- McHale, I., Morton, A., (2011). A Bradley-Terry Type Model for Forecasting Tennis Match Results. *International Journal of Forecasting*, 27, 619-630.
- Metcalfe, R., Burgess, S., Proud, S., (2011). Student Effort and Educational Attainment: Using the England Football Team to Identify the Education Production Function. Working paper.
- Mincer, J.,A., Zarnowitz, V., (1969). *The Evaluation of Economic Forecasts*, in “Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance”, pp. 1–46, *NBER*.
- Ottaviani, M., Sørensen, P., N., (2008), The Favorite-Longshot Bias: An Overview of the Main Explanations. *Handbook of Sports and Lottery Markets*, North Holland, edited by Hausch, D., B., Ziemba, W., T., 83-101.
- Parlour, C., A., Seppi, D., J., (2008). Limit Order Markets: A Survey. *Handbook of Financial Intermediation and Banking*, edited by Boot, A., W., A., Thakor, A., V., Elsevier.
- Shin, H., S., (1993). Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims. *The Economic Journal*, 103, 1141-1153.
- Spann, M., Skiera, B., (2009). Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters. *Journal of Forecasting*, 28, 55-72.
- Snowberg, E., Wolfers, J., (2010). Explaining the Favorite-Long Shot Bias: Is it Risk Love or Misperceptions? *Journal of Political Economy*, 118, 723-746.
- Surowiecki, J., (2005). *The Wisdom of Crowds: Why the Many are Smarter Than the Few*. Abacus, 3rd edition.
- Vaughan Williams, L., Reade, J., J., (2016). Forecasting Elections. *Journal of Forecasting*, 35, 308-328.

- Webbink, D., Van Erp, J., Van Gastel, F., (2016). The Effect of Media Exposure of Suspects on Solving Crime. *Economic Journal*, forthcoming.
- Weitzman, M., (1965). Utility Analysis and Group Behavior: An Empirical Study. *Journal of Political Economy*, 73, 18-26.

Figures and Tables

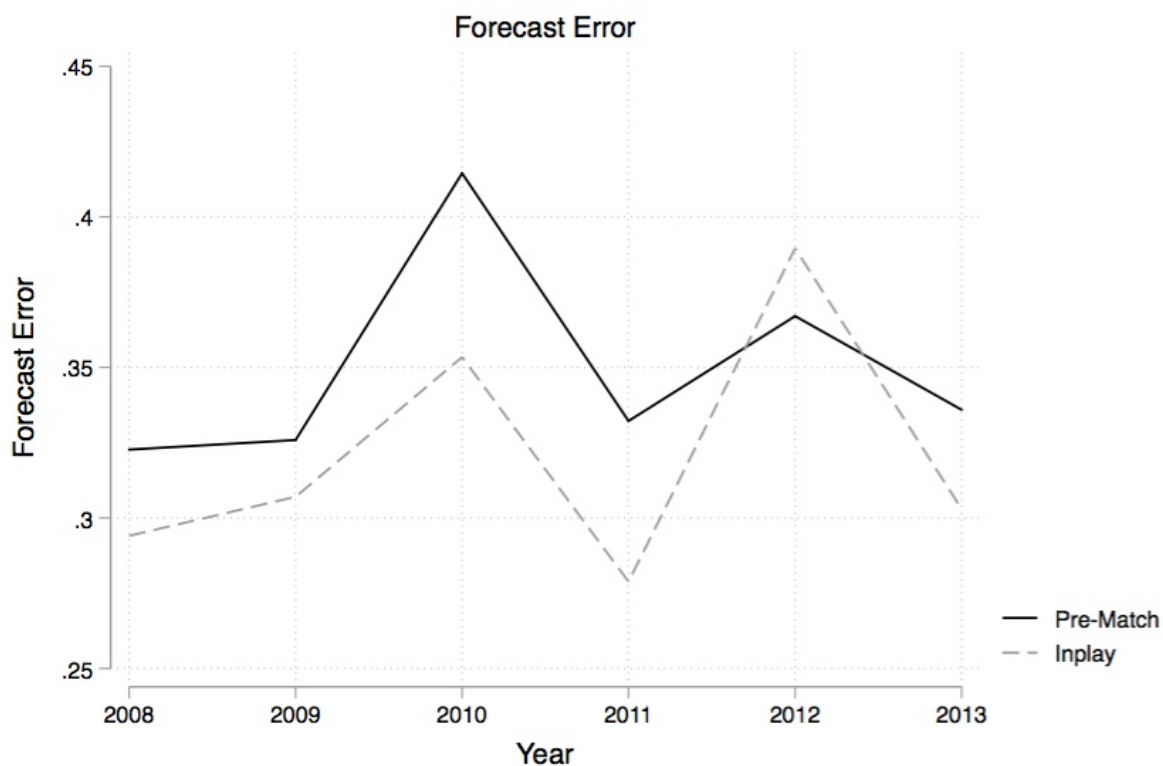


Figure 1: **Forecast Errors.** Plots of average forecast errors – defined as the absolute difference between the realized value of the bet (0/1) and the price (implied win probability) of the bet – between 2008 and 2013. In even years there was a clashing soccer tournament. Separate lines are plotted for pre-match and inplay betting.

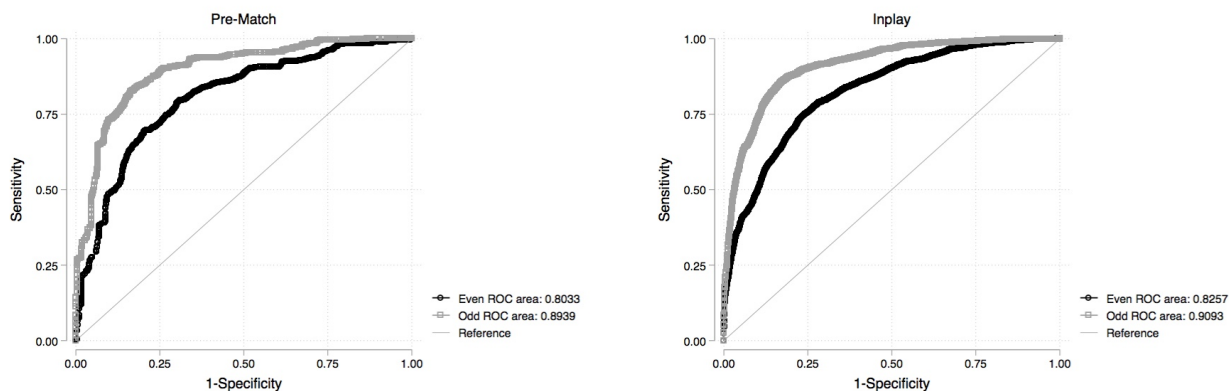


Figure 2: **ROC Curves.** Plots of Receiver Operating Characteristic (ROC) curves for odd years (no soccer tournament) and even years (clashing soccer tournament). Predictions are derived from the logit regression estimates in Table 6, and evaluated against actual win outcomes. Sensitivity, the proportion of match winners correctly identified by the model (and, by extension, the betting prices), is plotted against (1-Specificity), where Specificity is the proportion of losers correctly identified by the model. A larger area under the ROC curve implies superior forecasting of outcomes. Separate plots are produced for pre-match and inplay betting prices/forecasts.

Table 1: Internal Validity

Panel A: Baseline			
Variables	(1) Big Four	(2) Top Ten	(3) Ranking
Even Year	0.164 (0.498)	-0.0925 (0.423)	3.830 (6.220)
Constant	-2.853*** (0.158)	-2.039*** (0.0959)	82.98*** (2.186)
Observations	660	660	659
(Pseudo) R-squared	0.000829	0.000302	0.000
Panel B: With Time Drift			
Variables	(1) Big Four	(2) Top Ten	(3) Ranking
Even Year	-0.0894 (0.460)	-0.233 (0.353)	4.867 (5.972)
Year	-0.255* (0.140)	-0.140 (0.133)	1.043 (1.368)
Constant	-1.910*** (0.632)	-1.499*** (0.512)	78.82*** (6.513)
Observations	660	660	659
(Pseudo) R-squared	0.0211	0.00759	0.001

Robust standard errors – clustered at the year level – in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Regressions to establish the internal validity of the experiment. The three dependent variables are 1) an indicator variable equalling 1 if the player was a member of the ‘big four’, 2) an indicator variable equalling 1 if the player was in the top ten at the time of the tournament, and 3) the ranking of the player at the time of the tournament. These variables are regressed on an indicator variable equalling 1 if the match took place year in an even year (clashing soccer tournament) in Panel A. In Panel B we add the year – numbered 1-6 for 2008-2013 – to capture any drifts over time. Regressions (1) and (2) are logit regressions, while regression (3) is estimated using OLS. There are two observations for each match, one for each player.

Table 2: Summary Statistics

	(1)	(2)	(3)	(4)	(5)
	N	mean	sd	min	max
Pre-Match					
Trade	9,655,142	.01	.103	0	1
Trade Size	102,614	417	3,963	.007	739,165
Forecast Error	9,642,124	.351	.196	.012	.978
Inplay					
Trade	3,629,520	.162	.368	0	1
Trade Size	588,636	580	2,456	.007	276,455
Forecast Error	3,157,058	.322	.214	.001	.998

Summary statistics on all key measures used in the paper. Precise definitions of each of these measures can be found in Section 2.

Table 3: Trades				
Panel A: Baseline				
Variables	(1) Trade	(2) Trade	(3) Trade Size	(4) Trade Size
Even Year	-0.00220** (0.00107)	-0.0267* (0.0136)	-30.04 (69.88)	10.63 (52.52)
Constant	0.0120*** (0.000817)	0.176*** (0.0103)	430.3*** (49.38)	575.3*** (37.68)
Period	Pre-Match	Inplay	Pre-Match	Inplay
Observations	9,655,142	3,629,520	102,614	588,636
R-squared	0.000	0.001	0.000	0.000
Panel B: With Time Drift				
Variables	(1) Trade	(2) Trade	(3) Trade Size	(4) Trade Size
Even Year	-0.00246** (0.00103)	-0.0249* (0.0141)	-20.82 (72.67)	-4.378 (54.29)
Year	-0.000276 (0.000350)	0.00161 (0.00397)	8.642 (21.35)	-15.95 (14.30)
Constant	0.0131*** (0.00142)	0.169*** (0.0177)	397.3*** (91.18)	640.4*** (68.87)
Period	Pre-Match	Inplay	Pre-Match	Inplay
Observations	9,655,142	3,629,520	102,614	588,636
R-squared	0.000	0.001	0.000	0.000

Robust standard errors – clustered at the selection level – in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Regressions to establish whether the frequency of trades (captured by a trade indicator), and the size of trades (measured in GBP), varied across odd years (no soccer tournament) and even years (clashing soccer tournament). Separate regressions are run for pre-match and inplay betting. In Panel B we add the year – numbered 1-6 for 2008-2013 – to capture any drifts over time.

Table 4: Forecast Errors

Panel A: Baseline		
Variables	(1) Forecast Error	(2) Forecast Error
Even Year	0.0422** (0.0183)	0.0472*** (0.0155)
Constant	0.331*** (0.0123)	0.298*** (0.00972)
Period	Pre-Match	Inplay
Observations	9,642,124	3,157,058
R-squared	0.012	0.012
Panel B: With Time Drift		
Variables	(1) Forecast Error	(2) Forecast Error
Even Year	0.0485** (0.0190)	0.0593*** (0.0172)
Year	0.00672 (0.00517)	0.0106** (0.00451)
Constant	0.305*** (0.0224)	0.255*** (0.0214)
Period	Pre-Match	Inplay
Observations	9,642,124	3,157,058
R-squared	0.015	0.019
Robust standard errors – clustered at the selection level – in parentheses		
*** p<0.01, ** p<0.05, * p<0.1		

Regressions to establish whether forecast errors – defined as the absolute difference between the realized value of the bet (0/1) and the price (implied win probability) of the bet – varied across odd years (no soccer tournament) and even years (clashing soccer tournament). Separate regressions are run for pre-match and inplay betting. In Panel B we add the year – numbered 1-6 for 2008-2013 – to capture any drifts over time.

Table 5: Favourite-Longshot Bias

Variables	(1) Win	(2) Win	(3) Win	(4) Win
Implied Win Probability	1.358*** (0.0693)	1.079*** (0.0918)	1.247*** (0.0396)	1.038*** (0.0635)
Constant	-0.172*** (0.0423)	-0.0334 (0.0544)	-0.123*** (0.0329)	-0.0198 (0.0445)
Period	Pre-Match	Pre-Match	Inplay	Inplay
Year	Odd	Even	Odd	Even
Observations	4,921,850	4,720,274	1,579,722	1,577,336
R-squared	0.459	0.274	0.499	0.317

Robust standard errors – clustered at the selection level – in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Regressions to establish whether the degree of the favourite-longshot bias – measured by the deviation of the Implied Win Probability coefficient from 1 – varied across odd years (no soccer tournament) and even years (clashing soccer tournament). Separate regressions are run for odd and even years, and pre-match and inplay betting.

Variables	(1) Win	(2) Win	(3) Win	(4) Win
Implied Win Probability	8.774*** (1.150)	5.506*** (0.755)	8.348*** (0.725)	5.434*** (0.580)
Constant	-4.331*** (0.594)	-2.720*** (0.407)	-4.170*** (0.409)	-2.728*** (0.341)
Period	Pre-Match	Pre-Match	Inplay	Inplay
Year	Odd	Even	Odd	Even
Observations	4,921,850	4,720,274	1,579,722	1,577,336
Pseudo R-squared	0.409	0.221	0.449	0.259

Robust standard errors – clustered at the selection level – in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Logit regressions for predicting the winners and losers of Queen’s Club Tennis Championships matches. These predictions are then evaluated with ROC curves (see Figure 2). Separate regressions are run for odd and even years, and pre-match and inplay betting.