The wisdom of amateur crowds: evidence from an online community of sports tipsters

The Wisdom of Amateur Crowds: Evidence from an Online Community of Sports Tipsters*

Alasdair Brown†

University of East Anglia

J. James Reade‡

University of Reading

Institute for New Economic Thinking, Oxford

July 4, 2018

*We would like to thank the Editor, Robert Dyson, and four anonymous referees for very helpful comments and suggestions. We are also grateful to Peter Norman Sørensen for discussions on this topic. Part of 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

‡Department of Economics, University of Reading, Reading RG6 6AA, UK. E-mail: j.j.reade@reading.ac.uk

Abstract

We analyse the accuracy of crowd forecasts produced on Oddsportal, an online community of amateur sports tipsters. Tipsters in this community are ranked according to the betting return on their tips, but there are no prizes for accuracy. Nevertheless, we find that aggregated tips in this community contain information not in betting prices. A strategy of betting when a majority predict an outcome produces average returns of 1.317% for 68,339 events. The accuracy of these forecasts stems from the wisdom of the whole crowd, as selecting sections of the crowd based on experience or past forecast accuracy does not improve betting returns.

Keywords: OR in sports, sports betting, tipsters, wisdom of crowds

1 Introduction

Predicting the outcome of sporting events, particularly in a way unanticipated by bookmakers, is of prime interest to gamblers. Sports betting is estimated to be worth somewhere between 700 billion and 1 trillion worldwide per annum[1], which, even allowing for some exaggeration in that figure, clearly demonstrates that there are substantial sums at stake.

Traditionally, bettors may have decided to devise a model to forecast outcomes and see if these models produced information not in betting prices. Examples of academic work in this area are numerous and include Dixon and Coles (1997), Klaassen and Magnus (2003), Dixon and Pope (2004), Goddard and Asimakopoulos (2004), Easton and Uylangco (2010), and McHale and Morton (2011). More recently, and particularly after the publication of Surowiecki (2005) and the revival of the Galton (1907) 'wisdom of crowds' idea, there has been an interest in crowd-sourcing predictions. The wisdom of crowds operates on the premise that an averaging of forecasts eliminates individual prediction errors, and leads to greater accuracy.

Recent evidence suggests that there is indeed wisdom in the crowd when it comes to sports forecasting. Schumaker *et al.* (2016) and Brown *et al.* (2018) found that Twitter sentiment, or tone, could be leveraged to amass positive returns in English Premier League

[1] http://www.bbc.co.uk/sport/football/24354124

soccer betting. Peeters (2018) found that information from Transfermarkt valuations – where online users submit transfer valuations of soccer players – could be used to generate sizeable betting returns in matches.

In this paper we analyse predictions collected on Oddsportal, a betting comparison website which also hosts an online community of sports tipsters. Members of the Oddsportal community are ranked according to the betting return on their tips. The crowds on Oddsportal are smaller than Twitter, for example, but because of the ranking criteria these crowds are specifically tasked with identifying when betting markets are mispriced (i.e. when there is information not in betting prices). This setting therefore provides small, but highly-targeted, crowd-sourced predictions.

We set out to answer two questions. 1) Can Oddsportal tips be used to improve betting returns? And 2) does the informational content of these crowd-sourced predictions stem from the full crowd, or a subset of skilled or experienced individuals?

We find that Oddsportal tips can be used to generate positive betting returns. A strategy of betting when a majority of tipsters predict an outcome produces average returns of 1.317% for 68,339 events. This shows that even amateur crowds, with no tangible prizes for accurate forecasting, still produce information not in market prices. In further analysis we find that limiting the crowd to tipsters with more experience (more past tips) or more skill (higher historical returns on their tips) does not improve betting returns. This suggests that the accuracy of these crowd forecasts stems from the whole crowd, rather than just a select few tipsters.

2 Related Literature

Our work firstly contributes to the literature on sports tipsters. Deschamps and Gergaud (2005, 2008) and Forrest and Simmons (2000) found evidence that tipsters produced highly dispersed forecasts which ignored public information. The professional tipsters in Forrest and Simmons (2000) and Spann and Skiera (2009) performed poorly when considered against betting markets, but did improve these market forecasts when used in conjunction. For example, in Spann and Skiera (2009) the betting market predicted the correct winner 53.69% of the time in isolation, but 56.52% of the time when used in combination with aggregated tips. However, because of the large margins in the market considered, these combined forecasts did not produce positive returns (-9.08%). Reade (2014) considered the accuracy of Oddsportal soccer tips, but considered these tips as stand-alone forecasts rather than predictions of betting market mispricing.

Our paper also contributes to a literature on the wisdom of experts and laypeople (people without professional or specialist knowledge) in forecasting sports events. Experts outperformed laypeople in Pachur and Biele (2007), in part because laypeople forecasted based simply on name recognition (Goldstein and Gigerenzer, 2002). O'Leary (2017) found that a crowd of laypeople were more accurate than a smaller (n=5) group of experts, but did not evaluate whether these layperson predictions could produce positive betting returns. Herzog and Hertwig (2011) examined whether laypeople predictions could add to sports betting prices, and found not. Amateur tipsters on Oddsportal, on the other hand – who are not professionals but have self-selected themselves into predicting these events – produce information not in betting prices and, due to lower margins, can yield positive returns (1.317%).

Our paper is also related to a more general literature on the efficiency of betting prices, surveyed in Vaughan Williams (2005). For example, Ma *et al.* (2016) found that horse race betting markets failed to incorporate an important variable – the time since a horse last ran – into betting prices. Hwang and Kim (2015) found that betting market prices were poorly calibrated for extreme probabilities. The betting returns generated by Oddsportal predictions in our study are also indicative of market inefficiency, as information contemporaneously available to individuals and the crowd is not incorporated into betting prices.

This may be surprising to some, as Hayek (1945) argued that markets are well-suited to aggregate dispersed information, and the efficient market hypothesis of Fama (1970) may imply that markets should hold primacy in matters of forecasting. Part of the reason of the success of Oddsportal may be due to the payoff structure of forecasting contests. (Oddsportal, in effect, run an infinite-horizon forecasting contest with rankings determined by the betting returns on tips, albeit with no prizes). Pfeifer *et al.* (2014), Ottaviani and Sørensen (2005, 2006) and Lichtendahl *et al.* (2013) modelled forecasting contests and showed that forecasters will overweight their private information in a bid to win the contest. Put another way, the convexity of the prize schedule (e.g. winner-takes-all) encourages individuals to take risks and base their forecasts (solely) on their private information. As a result, individual forecasting errors may be large (as public information is ignored), but aggregated crowd forecasts will be more accurate as there is less repeated counting of public information ('public knowledge bias'). In markets, on the other hand, there is perhaps less incentive to ignore public information – as payoffs are not convex or dictated by relative rank – and therefore this may explain why forecasting contests can add information to that produced within markets.

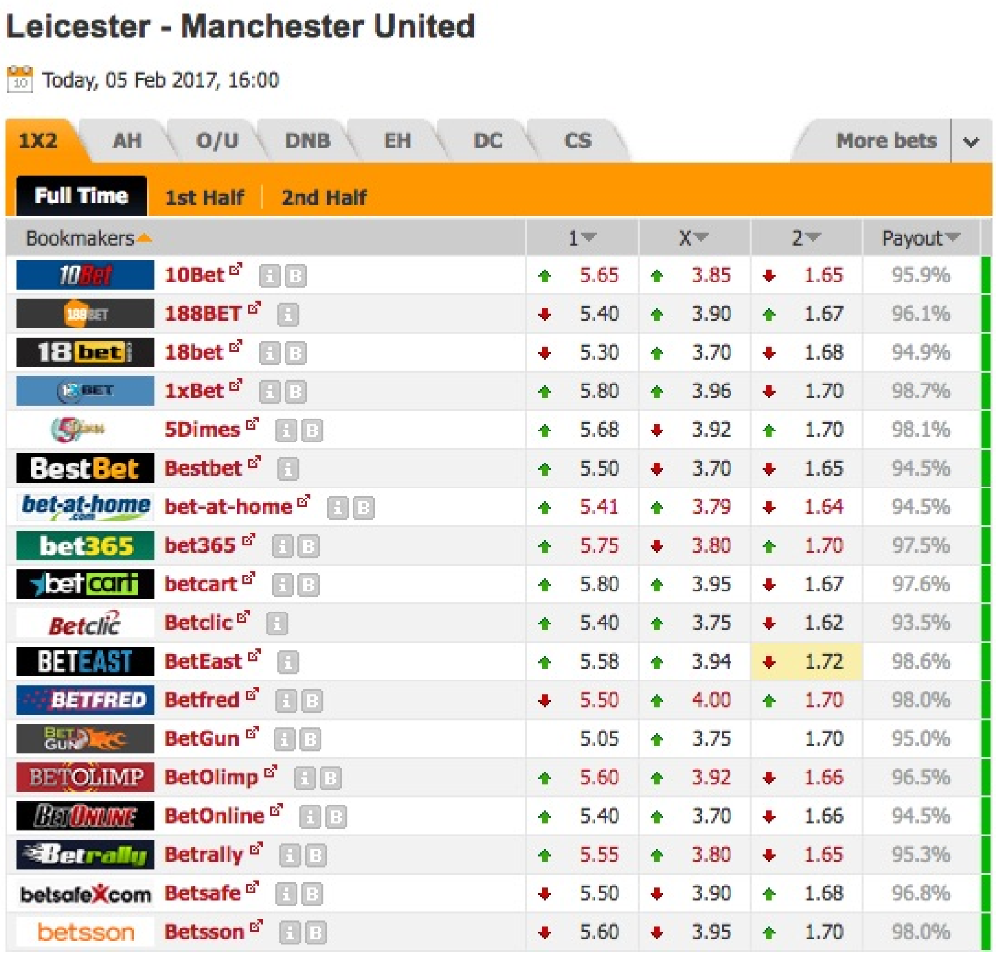
What is perhaps most striking about Oddsportal is that there are no tangible prizes – only the intangible esteem of ones' online peers – and yet there is still information contained in the forecasts made by these amateur tipsters.

Our results remind us of the findings in Servan-Schreiber et al. (2004), where play prediction markets performed as well as real-money prediction markets. Tipsters in our setting have weaker incentives than bookmakers and other participants in betting markets. Nevertheless, important information is produced in this low-stakes tipster community. This suggests that high-powered incentives are not the only consideration when generating accurate crowd forecasts.

In relation to the recent literature on crowd-sourced predictions of sporting events (e.g. Schumaker *et al.*, 2016), Brown *et al.*, 2018, and Peeters, 2018), the Oddsportal setting we study in this paper allows for two innovations. Firstly, we can examine whether targeted forecasts – on whether the bets are mispriced – can offset smaller crowd sizes and still produce profitable crowd forecasts. (It appears that they can). Secondly, as we have rich data on the full history of tipster predictions, we can analyse whether smaller crowds, made up of only the most experienced or skillful tipsters, can outperform the forecasts produced by the whole crowd. (It would appear not).

3 Data

The setting for our study is oddsportal.com, a website founded in 2008. The website serves two functions. Firstly it has an odds comparison function, providing the quoted odds from more than 80 bookmakers plus two betting exchanges, Betfair and Matchbook. The odds relate to 22 different sports from soccer (Association Football) to mixed martial arts. A screenshot of the odds comparison for the 5th February 2017 English Premier League match between Leicester City and Manchester United can be found in Figure 1. For illustration, we display only the first 18 bookmakers. The remaining bookmakers and the two betting exchanges were to be found underneath. In addition to the odds on the match outcome (home win/draw/away win), which we display, the website collates odds from the same bookmakers on the correct score and a range of other exotic bets.



Leicester - Manchester United

Today, 05 Feb 2017, 16:00

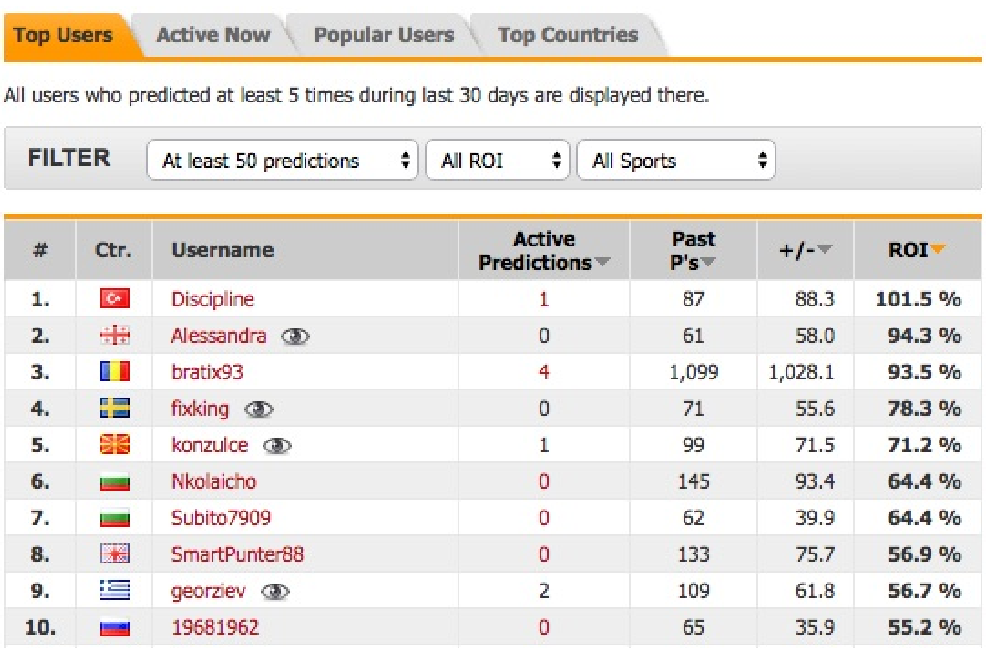
1X2 | AH | O/U | DNB | EH | DC | CS | More bets

Full Time | 1st Half | 2nd Half

Bookmakers	1	X	2	Payout
10Bet	5.65	3.85	1.65	95.9%
188BET	5.40	3.90	1.67	96.1%
18bet	5.30	3.70	1.68	94.9%
1xBet	5.80	3.96	1.70	98.7%
5Dimes	5.68	3.92	1.70	98.1%
Bestbet	5.50	3.70	1.65	94.5%
bet-at-home	5.41	3.79	1.64	94.5%
bet365	5.75	3.80	1.70	97.5%
betcart	5.80	3.95	1.67	97.6%
Betclic	5.40	3.75	1.62	93.5%
BetEast	5.58	3.94	1.72	98.6%
Betfred	5.50	4.00	1.70	98.0%
BetGun	5.05	3.75	1.70	95.0%
BetOlimp	5.60	3.92	1.66	96.5%
BetOnline	5.40	3.70	1.66	94.5%
Betrally	5.55	3.80	1.65	95.3%
Betsafe	5.50	3.90	1.68	96.8%
Betsson	5.60	3.95	1.70	98.0%

Figure 1: **Odds Comparison.** A screenshot of the odds comparison on oddsportal.com. The screenshot relates to the 5th February 2017 English Premier League match between Leicester City and Manchester United. Only the first 18 bookmakers are displayed, for illustration.

The second function of oddsportal.com is the hosting of an online community of sports tipsters. Registered users of the site can predict sporting outcomes, and they are then ranked according to the betting return on their tips. (The average bookmaker price is used when calculating the return). In Figure 2 we display a screenshot of the leaderboard of tipsters. There are various ways to filter this list, but in this particular case the tipsters are sorted by ROI (return on investment), and the list is restricted to those users with at least 50 historical predictions across all sports. Users may choose to keep their picks secret (with an eye indicating that this choice has been made), but the majority of users choose to share their picks with other users. This brings us to the Tips Feed (see Figure 3), where the most recent tips by all users (or just users you are following) are displayed. Users can comment on or 'Like' the tips made by other users, and this facility creates a social network feel to the website.



Top Users | Active Now | Popular Users | Top Countries

All users who predicted at least 5 times during last 30 days are displayed there.

FILTER | At least 50 predictions | All ROI | All Sports

#	Ctr.	Username	Active Predictions	Past P's	+/-	ROI
1.		Discipline	1	87	88.3	101.5 %
2.		Alessandra	0	61	58.0	94.3 %
3.		bratix93	4	1,099	1,028.1	93.5 %
4.		fixking	0	71	55.6	78.3 %
5.		konzulce	1	99	71.5	71.2 %
6.		Nkolaicho	0	145	93.4	64.4 %
7.		Subito7909	0	62	39.9	64.4 %
8.		SmartPunter88	0	133	75.7	56.9 %
9.		georziev	2	109	61.8	56.7 %
10.		19681962	0	65	35.9	55.2 %

Figure 2: **Tipster Rankings.** A screenshot of the rankings on oddsportal.com. In this screenshot tipsters are ranked by ROI (return on investment), which is the hypothetical betting return on their tips.

We collected the whole history of tips of 4,995 random tipsters on the website.[2] It is important that we sample tipsters randomly, because any sampling of the best or worst tipsters would of course bias our results. There are 1.79 million tips in our sample, relating to 231,842 different sporting events. Some of these tips relate to the main match outcomes,

[2]We collected the names of all user accounts on January 6th 2017, and then scraped information on users from that list at random, using the random.shuffle() function in the Python programming language to re-order the list (pseudo) randomly.

and some relate to more exotic bets. In total, there are 310,127 different bets subject to tips. The first tip in our sample was made on the 31st January 2012, and the last tip in our sample was made on the 17th January 2017, the last day of our sweep. For each of the bets involved in a tip we observe the average bookmaker odds on all outcomes. To ensure that each of our observations are independent in our later regressions, we focus only on the first outcome (e.g. home win or player 1 win). There are 535,527 tips on this outcome.

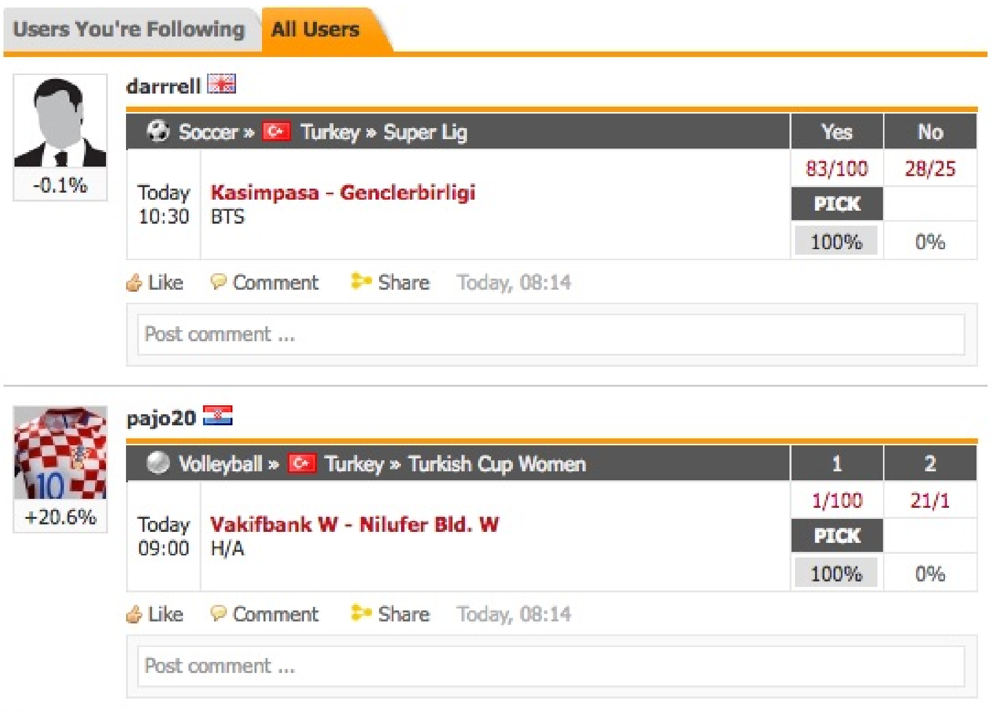


Figure 3: **Tips Feed.** A screenshot of the tips feed on oddsportal.com, which lists the latest tips. Users can choose to observe all tips in this feed, or tips only by tipsters that they are following.

In Table 1 we present summary statistics on all bets, and the bets picked out by tipsters. The average implied win probability – calculated as $1/Odds$ where the $Odds$ include the stake – is 0.477 on all bets on outcome 1. (This is higher than 0.33 as home wins are more frequent than draws or away wins). When we summarize tipster picks, however, we see an even higher average implied win probability of 0.545. Tipsters seem to select more favourites than longshots. In the third and fourth rows, we also look at experienced and skilled tipsters. These designations are made for each bet/event, and therefore experienced tipsters are defined as those who have previously lodged more tips than the median tipster who lodged a tip on the same event, and skilled tipsters are defined as those who have, at the time, a higher hypothetical return on their tips than the median tipster who lodged a tip on the same event. Experienced and skilled tipsters both pick more favourites than the average tipster.

Table 1: Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Implied Win Probability	N	mean	sd	min	max	skewness	kurtosis
All Bets	986,437	0.477	0.197	0.00281	0.990	0.029	2.4
Tipster Picks	535,527	0.545	0.155	0.00631	0.990	-0.185	3.287
Experienced Tipster Picks	219,325	0.546	0.147	0.00631	0.990	-0.213	3.4
Skilled Tipster Picks	217,049	0.549	0.147	0.00982	0.990	-0.197	3.37

Summary statistics on the implied win probability of all bets, bets picked out by tipsters, bets picked out by experienced tipsters, and bets picked out by skilled tipsters.

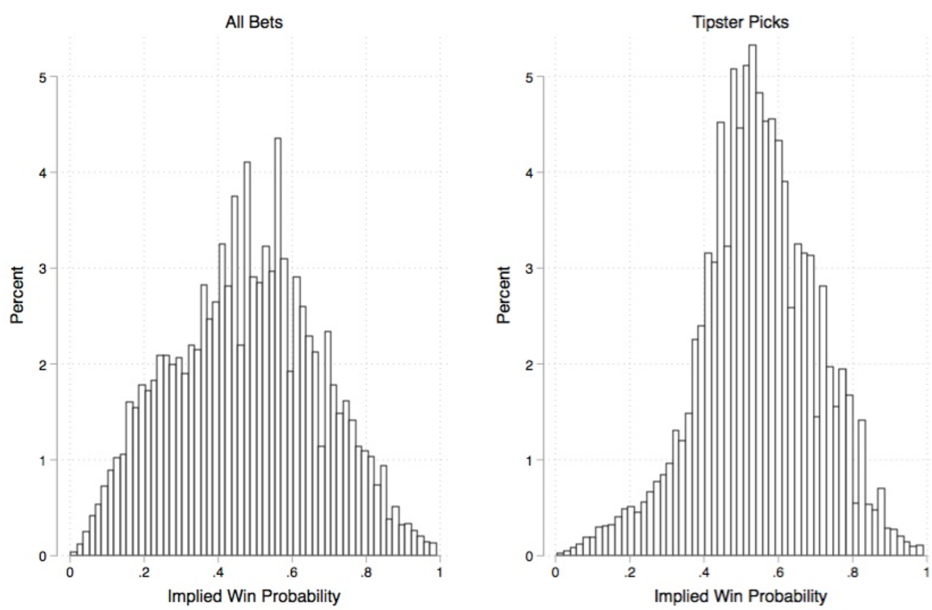


Figure 4: **Histogram: Tipster Picks.** Histograms of the implied win probability of all bets (left), and bets picked out by tipsters (right).

We summarise these patterns in Figures 4-6. We plot histograms of the implied win probability of all bets, all bets picked by a tipster, all bets picked by an experienced tipster, and all bets picked by a skilled tipster. There is a general shift rightward – toward the favourites – when we consider bets picked by tipsters, particularly those designated as experienced or skilled. There is also higher kurtosis, and more negative skewness, amongst the choices made by tipsters.

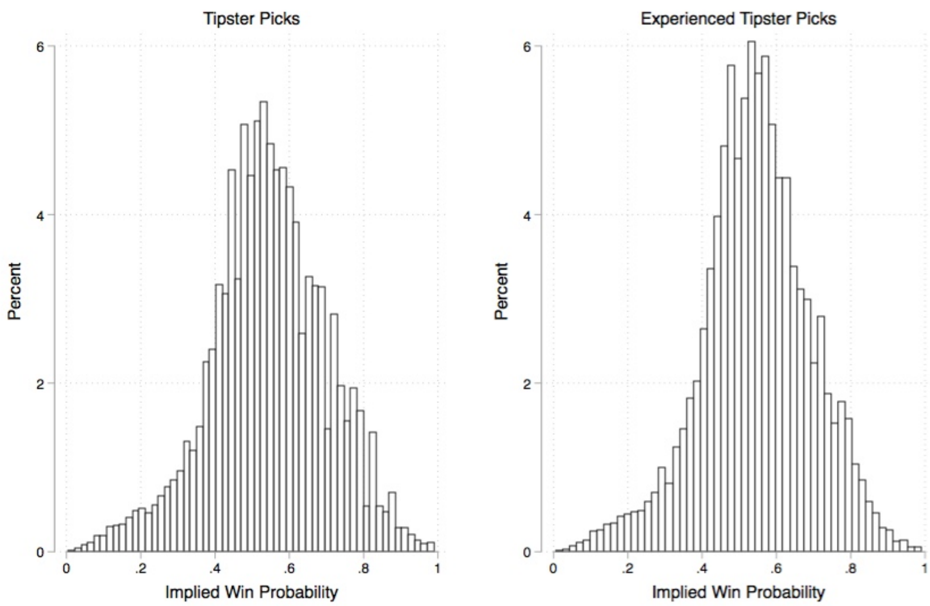


Figure 5: **Histogram: Experienced Tipster Picks.** Histograms of the implied win probability of bets picked out by tipsters (left), and bets picked out by experienced tipsters (right).

For our analysis in Section 4 we will be interested in collating these predictions. For example, we will analyse whether a team/individual is more likely to win (than betting prices suggest) if a majority of tipsters have predicted that outcome. Before we proceed to the analysis, it is worth briefly summarising the number of tips on each event/bet. The average event received 3.18 tips. The distribution of the number of tips is highly positively skewed, however, with one event subject to 144 tips from our random sample of tipsters, and more than 50% of events subject to only one tip.

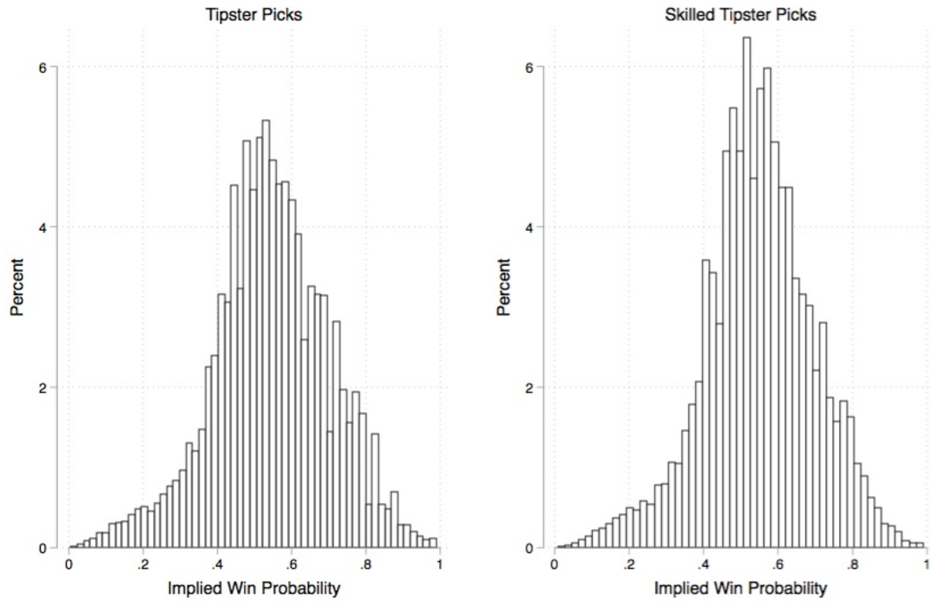


Figure 6: **Histogram: Skilled Tipster Picks.** Histograms of the implied win probability of bets picked out by tipsters (left), and bets picked out by skilled tipsters (right).

4 Analysis

For much of the analysis in this paper, we estimate a regression of the following form:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i \tag{1}$$

y_i is an indicator variable, equalling 1 if outcome i occurred (i.e. the player/team won) and 0 otherwise (the player/team lost or drew), x_i is the implied probability of outcome i as measured from the average bookmaker odds, z_i is a variable capturing some element of the aggregated tipster forecasts, and ϵ_i is an error term. A variant of this equation, without the z_i term, is commonly referred to as the Mincer-Zarnowitz regression (Mincer and Zarnowitz, 1969). The idea is to examine whether tipster forecasts add any information to that already embedded in betting prices. To ensure the independence of our observations, we include only the tips and bets on outcome 1 for each bet/event. The standard errors for all of our upcoming regressions are heteroskedasticity-consistent. We estimate Equation (1) by OLS (Ordinary Least Squares), but all of our results are qualitatively similar with a binary logit model. Throughout the results section we will stick with convention and indicate significance at the 1%, 5% and 10% levels using ***, **, and * respectively. However, given the number of observations in our data we will not assign much weight to findings with significance at either the 5% or 10% levels.

We present our first results in Table 2. Before we incorporate tipster predictions in our regression model, we first estimate the Mincer-Zarnowitz regression to establish the general efficiency of this betting market, and specifically ascertain whether there is a favourite-longshot bias. The favourite-longshot bias is the long-standing regularity – dating at least from Griffith (1949) – that returns differ between favourites and longshots. More often than not, returns have been found to be higher for bets on favourites. Based on the results of our first regression, however, there is actually a small and significant negative bias in this market, as the estimate of β_1 is less than 1.[3] This is not uncommon, as Busche and Hall (1988) found a similar direction to the bias in Hong Kong horse racing data. Nevertheless, as Oddsportal tipsters disproportionately back favourites, this might decrease the returns available to strategies based on these tips.

[3] In all of the following regressions, bar one, the estimate of β_1 is statistically distinct from 1 at the 5% level.

Table 2: Main Results

	(1)	(2)	(3)	(4)
Variables	Win	Win	Win	Win
Implied Win Probability	0.991***	0.983***	0.983***	0.983***
	(0.00332)	(0.00339)	(0.00339)	(0.00339)
Proportion of Tipsters		0.0193***		0.0202***
		(0.00185)		(0.00536)
Tipster Majority			0.0161***	-0.000918
			(0.00168)	(0.00487)
Constant	-0.00375**	-0.00915***	-0.00732***	-0.00922***
	(0.00184)	(0.00191)	(0.00188)	(0.00193)
Observations	310,127	310,127	310,127	310,127
R-squared	0.174	0.174	0.174	0.174
Out-of-Sample MSFE	0.2050588	0.205004	0.2050166	.2049997

Robust standard errors in parentheses

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

Regressions to establish whether an aggregation of tips from Oddsportal can predict outcomes after controlling for betting market prices. In regression 1, an indicator variable equalling 1 if the bet won, is regressed on the implied win probability as calculated from the betting odds. In regressions 2 and 3 respectively, we separately add the proportion of tipsters that predicted that outcome, and an indicator variable equalling 1 if a majority of tipsters predicted that outcome. In regression 4 we use all three regressors. In the bottom row we present out-of-sample mean squared forecast errors (MSFE) for each of the four models, based on estimates from the first half of the data (January 31st 2012 to 11.30 GMT April 12th 2015) which are used to generate predictions with the second half of the data (12.30 GMT April 12th 2015 to 17th January 2017).

In the second column of Table 2 we run the same regression, but this time add the proportion of tipsters that predicted the outcome as an additional regressor. For example, if 5 out of 7 tipsters back team A to win, this proportion is 0.714. We find that this is a significant predictor of sporting outcomes. The greater the proportion of tipsters backing an outcome, the more likely it is to occur, even after controlling for betting prices. Then, in a third regression we use an indicator variable equalling 1 if a majority of tipsters (more than 50%) predicted the outcome and 0 otherwise. This is a coarser measure of tipster predictions, but is still a significant predictor of outcomes. An outcome is 1.61% more likely to occur if a majority of tipsters forecasted that outcome compared to if a majority did not forecast the outcome, after controlling for betting prices. In short, we find that an aggregation of Oddsportal tips contains information not found in betting market prices.

In our final regression in Table 2, we include the proportions of tipsters backing an outcome, and the indicator variable equalling 1 if a majority back the outcome, as regressors in the same regression. This allows us to examine whether there is a jump in the win probability of teams/individuals if a majority of tipsters predict the outcome. Based on our results in regression 4, this does not appear to be the case, as the majority indicator is not statistically significant. The information contained in aggregate tips seems to be a reasonably smooth function of the proportion of tipsters that back an event.

In addition to our main in-sample forecasting results, we also tested whether models which use information on tips performed better out-of-sample. To do this, we estimated the coefficients for all of the regressions in Table 2 for events in the first half of the data from January 31st 2012 to 11.30 GMT April 12th 2015. We then used these estimates to predict outcomes for events in the second half from 12.30 GMT April 12th 2015 to 17th January 2017. We then calculated mean squared forecast errors for each of the four models based on the second half of the data. These mean squared forecast errors are presented in the bottom row of Table 2. We find that all three models which incorporated information on tips (regressions 2 to 4) perform significantly better (at the 1% level) out-of-sample than the model based purely on betting prices (regression 1), as testified by lower mean squared forecast errors.

Next, we examine whether it is worth selecting a subset of tipsters and focusing only on their tips. To be specific, is there more information contained in the picks of experienced or skilled tipsters? We might expect that a tipster's prior record can guide us as to the informational content of their predictions. In different settings, Mannes *et al.* (2014) and Budescu and Chen (2015) found that forecasts can be improved by sorting on prior accuracy. Davis-Stober *et al.* (2015), on the other hand, show theoretically that crowd accuracy is a trade-off between diversity of opinions and forecasting ability (revealed in prior forecasts).

In Table 3 we begin by repeating regression 3 of Table 2, but this time looking only at events where there was more than one tip. (This will allow us to later divide up this crowd by experience or skill). An outcome is 1.8% more likely to occur if a majority of the crowd tipped the outcome compared to if a majority did not tip the outcome, after controlling for betting prices. In regressions 2 and 3, we then regress the win indicator variable on the implied win probability from the odds, and an indicator variable equalling 1 a majority of experienced/skilled tipsters backed that outcome.[4] Surprisingly, we find little benefit to relying on experienced or skilled tipsters. For example, an outcome is 1.45% more likely to occur if a majority of experienced tipsters predicted the outcome, and is 1.74% more likely to occur if a majority of skilled tipsters predicted the outcome. This compares unfavourably with our first results in Table 3, where we found that an outcome is 1.8% more likely to occur if a majority of all tipsters predicted the outcome. In short, by ignoring supposedly inexperienced or unskilled tipsters we do not improve the wisdom of the crowd. It should be said, however, that in all cases – 1) full crowd versus experienced crowd, 2) full crowd versus skilled crowd, and 3) skilled crowd versus experienced crowd – the coefficients in the three regressions are not significantly different at the 10% level. This means that the full crowd is no worse than the skilled crowd, but is also no better.[5]

[4]There is a high, but less than perfect, correlation of 0.7402 between a majority of all tipsters backing an outcome and a majority of experienced tipsters backing an outcome, and a similar correlation of 0.7458 between a majority of all tipsters backing an outcome and a majority of skilled tipsters backing an outcome.

[5]There were a very small number of cases where all the tipsters for an event had the same level of experience (number of past tips) or skill (return on prior tips), and therefore the indicator variables in regressions 2 or 3 of Table 3 were classified as missing. This explains the small differences in the number of observations in regressions 1 to 3.

Table 3: Sorting Tipsters According to Experience/Skill

	(1)	(2)	(3)
Variables	Win	Win	Win
Implied Win Probability	0.986***	0.991***	0.988***
	(0.00576)	(0.00570)	(0.00576)
Tipster Majority	0.0180***		
	(0.00259)		
Experienced Tipster Majority		0.0145***	
		(0.00254)	
Skilled Tipster Majority			0.0174***
			(0.00257)
Constant	-0.00454	-0.00569**	-0.00574**
	(0.00283)	(0.00284)	(0.00284)
Sample	Tips>1	Tips>1	Tips>1
Observations	147,439	147,399	147,322
R-squared	0.150	0.150	0.150

Robust standard errors in parentheses

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

Regressions to establish whether an aggregation of tips from Oddsportal can predict outcomes after controlling for betting market prices. This time we look separately at the full crowd (when the number of tips is greater than 1), experienced tipsters and skilled tipsters. In the regressions, an indicator variable equalling 1 if the bet won, is regressed on the implied win probability as calculated from the betting odds, and an indicator variable equalling 1 if a majority of (all/experienced/skilled) tipsters predicted that outcome.

In our final and most important analysis, we examine the returns available to betting strategies predicated on tipster forecasts. These results are presented in Table 4. We begin by summarising the returns to all bets. This gives us an idea of the transaction costs involved in betting. Average returns are -2.278% which, while negative, are still quite high by betting standards. We then look at the returns to betting when there is a tipster majority backing an outcome. Here the returns are marginally positive (0.0059%). Reflecting our earlier findings, the highest returns are available when following crowd predictions. Bets with a tipster majority – when the event was subject to more than 1 tip – return 1.317% on average. Finally, relying on experienced or skilled tipsters appears to be slightly detrimental (but not significantly so), returning 0.824% and 1.117% respectively. In short, there are higher returns available for following the whole crowd.

Table 4: Returns to Betting Strategies

	(1)	(2)	(3)	(4)	(5)
Returns (%)	N	mean	sd	min	max
All Bets	310,127	-2.278	127.0	-100	6,900
Tipster Majority	142,875	0.00594	108.5	-100	2,500
Tipster Majority, Tips>1	68,339	1.317	99.45	-100	1,800
Experienced Tipster Majority, Tips>1	74,013	0.824	103.6	-100	1,800
Skilled Tipster Majority, Tips>1	74,028	1.117	103.0	-100	1,800

A summary of percentage returns to various betting strategies are displayed. These are returns for all bets, returns for bets when a majority of tipsters predicted the outcome, returns for bets when a majority of tipsters predicted the outcome when there was more than 1 tip (the crowd), returns for bets when a majority of experienced tipsters predicted the outcome, and returns for bets when a majority of skilled tipsters predicted the outcome.

The returns to these betting strategies may be considered a lower bound. Oddsportal display the average betting odds at the time the tip was made, but a gambler placing a bet is likely to place bets at the highest odds offered, not least because Oddsportal provide a price comparison service precisely to identify the best odds. However, our decision to focus on the average rather than best odds is in part practical. When Oddsportal archive past tips, they only retain information on the average odds at the time of the tip, not the best odds. Moreover, even if we could observe the best odds, it is probably prudent to use the average odds. Bookmakers quoting outlying odds may refuse to accept large stakes, and have the right to cancel a bet before the event if they deem the odds to have been unreasonable. By using the average odds, we provide a conservative estimate of the wisdom of this amateur crowd.

5 Conclusion

We analyse whether crowd predictions from Oddsportal – an online community of sports tipsters – provide information that can be used to generate positive betting returns. Crowds on Oddsportal are often small, and comprised only of amateurs. However, these small crowds of amateurs provide highly-targeted predictions of betting market mispricing.

We find that following a strategy of betting when a majority of tipsters forecast an outcome yields average returns of 1.317% for 68,339 events. The accuracy of these forecasts seems to stem from the whole crowd, rather than a select few, as limiting the crowd on the basis of prior experience or prior forecast accuracy actually leads to slightly lower returns. This suggests, in line with Davis-Stober *et al.* (2015), that diversity of opinion is perhaps as important as forecasting pedigree when harnessing the wisdom of crowds.

One issue with the Oddsportal community design is that tipsters can observe the tips made by their predecessors. This presents an issue for researchers as we cannot disentangle individuals' beliefs from the crowd's beliefs. This also presents an issue for the use of such a platform for forecasting. Tipsters may be herding behind the tips of their predecessors, which would lead in some cases to correlated forecast errors and an inferior crowd forecast. It is possible that with a blind tipping system such correlated forecast errors would be reduced, and the returns to betting strategies predicated on amateur tipster crowd wisdom would be higher.

References

- Brown, A., Rambaccussing, D., Reade, J., J., Rossi, G., (2018). Forecasting with Social Media: Evidence from Tweets on Soccer Matches. *Economic Inquiry*, 56, 1748-1763.

- Budescu, D., V., Chen, E., (2015). Identifying Expertise to Extract the Wisdom of Crowds. *Management Science*, 61, 267-280.

- Busche, K., Hall, C., D., (1988). An Exception to the Risk Preference Anomaly. *Journal of Business*, 61, 337-346.

- Davis-Stober, C., P., Budescu, D., V., Broomell, S., B., Dana, J., (2015). The Composition of Optimally Wise Crowds. *Decision Analysis*, 12, 130-143.
- Deschamps, B., Gergaud, O., (2005). Strategic Forecasting in Rank-Order Tournaments: Evidence from Horse-Racing Tipsters. Working paper.
- Deschamps, B., Gergaud, O., (2008). Efficiency in Horse Race Betting Markets: The Role of Professional Tipsters, *Handbook of Sports and Lottery Markets*, North Holland, edited by Hausch, D., B., Ziemba, W., T., 341-354.
- Dixon, M. J., Coles, S., G., (1997). Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48, 265-280.
- Dixon, M., J., Pope, P., F., (2004). The Value of Statistical Forecasts in the UK Association Football Betting Market. *International Journal of Forecasting*, 20, 697-711.
- Easton, S., Uylangco, K., (2010). Forecasting Outcomes in Tennis Matches Using Within-Match Betting Markets. *International Journal of Forecasting*, 26, 564-575.
- Fama, E., F., (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, 25, 383-417.
- Forrest D., Simmons, R., (2000). Forecasting Sport: The Behaviour and Performance of Football Tipsters. *International Journal of Forecasting*, 16, 317-331.
- Galton, F., (1907). Vox Populi, *Nature*, 1949, 450-451.
- Goddard, J., Asimakopoulos, I., (2004). Forecasting Football Results and the Efficiency of Fixed-Odds Betting. *International Journal of Forecasting*, 23, 51-66.
- Goldstein, D., G., Gigerenzer, G., (2002). Models of Ecological Rationality: The Recognition Heuristic. *Psychological Review*, 109, 75-90.
- Griffith, R., M., (1949). Odds Adjustments by American Horse-Race Bettors. *The American Journal of Psychology*, 62, 290-294.

- Hayek, F., A., (1945). The Use of Knowledge in Society. *American Economic Review*, 35, 519-530.

- Herzog, S., M., Hertwig, R., (2011). The Wisdom of Ignorant Crowds: Predicting Sport Outcomes by Mere Recognition. *Judgement and Decision Making*, 6, 58-72.

- Hwang, J., H., Kim, M., S., (2015). Misunderstanding of the Binomial Distribution, Market Inefficiency, and Learning Behavior: Evidence from An Exotic Sports Betting Market. *European Journal of Operational Research*, 243, 333-344.

- Klaassen, F., J., G., M., Magnus, J., R., (2003). Forecasting the Winner of a Tennis Match. *European Journal of Operational Research*, 148, 257-267.

- Lichtendahl, K., C., Jr., Grushka-Cockayne, Y., Pfeifer, P., E., (2013). The Wisdom of Competitive Crowds. *Operations Research*, 61, 1383-1398.

- Ma, T., Tang, L., McGroarty, F., Sung, M., C., Johnson, J., E., (2016). Time is Money: Costing the Impact of Duration Misperception in Market Prices. *European Journal of Operational Research*, 255, 397-410.

- Mannes, A., E., Soll, J., B., Larrick, R., P., (2014). The Wisdom of Select Crowds. *Journal of Personality and Social Psychology*, 107, 276-299.

- McHale, I., Morton, A., (2011). A Bradley-Terry Type Model for Forecasting Tennis Match Results. *International Journal of Forecasting*, 27, 619-630.

- Mincer, J., A., Zarnowitz, V., (1969). The Evaluation of Economic Forecasts, in *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, 1-46, NBER.

- Ottaviani, M., Sørensen, P., N., (2005). Forecasting and Rank-Order Contests. Working paper.

- Ottaviani, M., Sørensen, P., N., (2006). The Strategy of Professional Forecasting. *Journal of Financial Economics*, 81, 441-466.

- O'Leary, D., E., (2017). Crowd Performance in Prediction of the World Cup 2014. *European Journal of Operational Research*, 260, 715-724.
- Pachur, T., Biele, G., (2007). Forecasting from Ignorance: The Use and Usefulness of Recognition in Lay Predictions of Sports Events. *Acta Psychologica*, 125, 99-116.
- Peeters, T., (2018). Testing the Wisdom of Crowds in the Field: Transfermarkt Valuations and International Soccer Results. *International Journal of Forecasting*, 34, 17-29.
- Pfeifer, P., E., Grushka-Cockayne, Y., Lichtendahl K., C., Jr., (2014). The Promise of Prediction Contests. *The American Statistician*, 68, 264-270.
- Reade, J., J., (2014). Information and Predictability: Bookmakers, Prediction Markets and Tipsters as Forecasters. *Journal of Prediction Markets*, 8, 43-76.
- Schumaker, R., P., Jarmoszko, A., T., Labedz Jr., C., S., (2016). Predicting Wins and Spread in the Premier League Using a Sentiment Analysis of Twitter. *Decision Support Systems*, 88, 76-84.
- Servan-Schreiber, E., Wolfers, J., Pennock, D., M., Galebach, B., (2004). Prediction Markets: Does Money Matter? *Electronic Markets*, 14, 243-251.
- 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.
- Surowiecki, J., (2005). *The Wisdom of Crowds: Why the Many are Smarter Than the Few*. Abacus, 3rd edition.
- Vaughan Williams, L., (2005). *Information Efficiency in Financial and Betting Markets*, Cambridge University Press.