Betting on a buzz, mispricing and inefficiency in online sportsbooks

Philip Ramirez J. James Reade Carl Singleton*

May 2021

Link to latest version

Abstract

Bookmakers sell claims to bettors that depend on the outcomes of professional sports events. Like other financial assets, the wisdom of crowds could help sellers to price these claims more efficiently. We use the Wikipedia profile page views of professional tennis players involved in over ten thousand singles matches to construct a buzz factor. This measures the difference between players in their pre-match page views relative to the usual number of views they received over the previous year. The buzz factor significantly predicts mispricing by bookmakers. Using this fact to forecast match outcomes, we demonstrate that a strategy of betting on players who received more pre-match buzz than their opponents can generate substantial profits. These results imply that sportsbooks could price outcomes more efficiently by listening to the buzz.

Keywords: Wisdom of crowds, Betting markets, Efficient Market Hypothesis, Forecast efficiency, Professional tennis
JEL codes: G14, G41, L83

*p.ramirez@pgr.reading.ac.uk, j.j.reade@reading.ac.uk and c.a.singleton@reading.ac.uk (corresponding author), Department of Economics, University of Reading, Whiteknights Campus, RG6 6EL, UK.

We would like to thank Giovanni Angelini, Luca De Angelis and Tho Pham for helpful comments.

1 Introduction

The size and ubiquity of online sports betting markets continues to increase. Most notably in recent years, the world's most successful online sportsbooks entered the U.S. after a 2018 Supreme Court ruling allowed states to legalise gambling at their own discretion.[1] As online sports betting markets have grown and replaced more traditional forms of gambling, lower transaction costs have increased competition and driven down bookmaker profit margins (i.e., the overround or vig) (Forrest, 2008). Over the same period, the amount of online information that bettors can use to form expectations about sports outcomes has increased. This includes detailed historical data about the participants and the setting of an event, the commentary and predictions of sports pundits and tipsters, and the so-called 'wisdom of crowds'. This latter term is used widely to describe instances where information aggregated from the decisions of many individuals improves forecasting and decision-making processes, compared with relying on a small number of expert positions (Galton, 1907; Surowiecki, 2004). Given the small profit margins and competition with the crowd-based betting exchanges (prediction markets), odds-setters must forecast outcomes and price the claims they sell to bettors more efficiently than ever before. It is natural to ask whether bookmakers are doing this successfully. In this paper, we use a specific practical example to demonstrate how online sportsbooks are vulnerable to information that could represent the wisdom of crowds.

Wikipedia, the free online encyclopedia, is an example of crowd wisdom. It has become the go-to online place for information about almost anything, including the the characteristics and form of sports people. We use this fact to construct what we call the Wiki Relative Buzz Factor, for over ten thousand Women's Tennis Association (WTA) singles matches since the beginning of the 2015 season.[2] These matches were all at the elite level of the sport and

[1] See Murphy v. National Collegiate Athletic Association, No. 16-476, 584 U.S. (2018), which ruled that the Professional and Amateur Sports Protection Act of 1992 was unconstitutional. As of 1 April, 2021, 11 states have legalised online sports betting: California, Delaware, Illinois, Indiana, Michigan, Nevada, New Hampshire, New Jersey, Pennsylvania, Rhode Island and West Virginia.

[2] We have no particular rationale for focusing on this sport and the women's game only. It is, however, convenient that odds on all these events were offered by a large number of online sportsbooks. Further, we

include the four annual Grand Slam tournaments. The buzz factor uses the numbers of page views on the Wikipedia profiles of players before their matches began. It is relative because it compares across players within a match. We call it buzz because it uses the profile page views on the day before a match relative to the typical numbers over the past 12 months. We then adapt the Mincer and Zarnowitz (1969) forecast evaluation framework, showing that the Wiki Relative Buzz Factor can significantly predict the systematic mispricing of bookmaker odds, with the higher buzz player being under priced. There is no significant evidence of a favourite or longshot bias in these markets, but bookmakers tended to significantly underprice a player who was substantially lower ranked than their opponent. Taking these results together, we can reject a sufficient condition for weak form market efficiency. To prove that these markets are inefficient, we generate probability forecasts of tennis match results by using the same model that detected the mispricing. Combining these forecasts with the Kelly criterion, which can be motivated from expected utility theory, we demonstrate substantial and sustained profits from exploiting the information contained in the Wiki Relative Buzz Factor. Specifically, we found a potential return on investment of 17-29% from applying the forecasting model at Bet365, the world's highest revenue online sportsbook, over five thousand potential bets on WTA matches between the beginning of the 2019 season and March of 2020. In contrast, using probability forecasts from the widely used Elo (1978) rating systems and the Kelly criterion would have generated substantial losses over the same samples of matches.

These results contribute to the growing literature attempting to elicit the value of crowd wisdom from the field and using this to test the Efficient Markets Hypothesis (Fama, 1965, 1970). Relevant to our study of betting markets, research has demonstrated how information from social media can predict what happens in financial markets, including cross-sectional stock returns (e.g., Avery et al., 2016; Chen et al., 2014; Sprenger et al., 2014) and the price movements of cryptocurrencies (Kraaijeveld and De Smedt, 2020). In a closely related study, Brown et al. (2018) discovered that the aggregate tone extracted from large numbers of Twitter

had built a dataset containing information about these events other research projects before using it to explore the questions in this paper.

posts contained significant information not in live betting exchange prices during football matches, especially in the aftermath of major events such as goals or red cards. Using a crowd explicitly making predictions, Brown and Reade (2019) found that the aggregated content from a community of online sports tipsters also contained information not in betting prices. Betting when the majority of the community predicted a particular outcome generated a small average positive return. Peeters (2018) also found that a crowd of sports fans could improve forecasting accuracy and generate profitable opportunities on betting markets. Specifically, forecasts based on the football player transfer market values on *transfermarkt.de*, and the implied strengths of international teams, proved more accurate than other standard predictors of match results, such as official team rankings or form-based rating systems.

This paper contributes more generally to the literature on the efficiency of betting and prediction markets, specifically for sports, much of which has focused on the favourite-longshot bias (for reviews see Vaughan Williams, 1999, Ottaviani and Sørensen, 2008 or Newall and Cortis, 2021). There is a small literature focused on the efficiency of tennis match betting markets (Abinzano et al., 2016, 2019; Forrest and Mchale, 2007; Lahvička, 2014; Štefan Lyócsa and Výrost, 2018). This literature has tended to find evidence of a longshot bias that is not large enough to overcome the bookmaker profit margin and prove inefficiency. The present paper also contributes to the use of professional sports to learn something about the practice of forecasting, in particular to some studies that have focused on professional tennis (e.g., Angelini et al., 2021a; Candila and Scognamillo, 2018; del Corral and Prieto-Rodríguez, 2010; Easton and Uylangco, 2010; Kovalchik and Reid, 2019; Kovalchik, 2020; McHale and Morton, 2011; Scheibehenne and Broder, 2007). The forecasting models introduced by these studies cannot normally outperform bookmakers without shopping around to find the best available odds (Angelini et al., 2021a; Kovalchik, 2016).

The rest of the paper proceeds as follows: Section 2 describes our dataset, a model to detect mispricing, and a simple betting strategy to test market efficiency using the model; Section 3 presents the results; and Section 4 concludes.

2 Data & Method

We collected information from tennis-data.co.uk for all WTA match results from the main draws of all tournaments, including the Grand Slams, between 1 January, 2015, and 16 February, 2020. This information includes the identity of players and tournaments, as well as when (local date) and where matches took place.[3] The dataset represents 10,522 matches, 443 players and 271 tournaments. It includes the WTA world rankings of the players immediately before each match, which are based on performances over the preceding year and are updated after a tournament is completed. We used the python packages *geopy* and *timezonefinder* to locate the coordinates of each city in the dataset and the time zones for each match location.

The main draw for a WTA tournament normally takes place a few days before the first round begins, after any qualification matches. All tournaments are in a knock-out format and the draw is seeded, except the end-of-season WTA Tour finals which have a round-robin stage. The seeding is based on world rankings going into a tournament. The average length of a WTA tennis match is around two and a half hours. A player can normally expect one to three days of rest between matches in a tournament. The lineup for a match is usually known at least the day before it starts, either after the first round draw or the completion of players' previous matches in the tournament, at which point betting odds will become available.

We collected betting odds from oddsportal.com for the winner and loser of a match at the time it began. In what follows, we generally use the average odds from the forty to sixty online bookmakers (sportsbooks) that were posted for any given match on oddsportal.com. We also use the highest (or best) available odds from the bookmaker sample for each match, as well as the specific odds from Bet365, the largest online bookmaker in the world by revenue, which offered odds on almost every match in the dataset.

[3]The local date gives the match start, which is important since matches can be played over multiple days due to stoppages, for example, due to the weather.

2.1 The Wikipedia relative buzz factor

To construct a measure of the pre-match buzz about the players, we collected daily (Coordinated Universal Time, UTC) Wikipedia page views of their English language profiles using the Pageview Application Programming Interface (API), a tool used to query the Wikipedia Foundation pageview data. A small number of observations in the WTA match dataset use maiden names, nicknames, or variations of abbreviations. Therefore, we were careful to ensure every player in the WTA dataset was matched to their Wikipedia profile page views. The mean number of page views for players on the day before a match took place was 1,079, with a median of 139, a standard deviation of 6,823 and a maximum of 429,245 (for Naomi Osaka, 7 September, 2018, the day before she won the US Open final and her opponent, Serena Williams, accused the umpire of being a "thief"). Panel (a) of Figure 1 shows kernel density plots of the log profile page views of players the day before a match took place. The distribution for match winners is generally to the right of that for match losers, suggesting that players with higher levels of interest in their profiles before a match were more likely to win. Panel (b) of Figure 1 shows the tighter distributions of the log median page views in the past year before a match, though with greater differences between the winner and loser distributions than in panel (a), suggesting the typical past number of profile page views could be a better predictor of subsequent success in a match.

To generate our 'Wiki Relative Buzz Factor' for each player-match observation in the dataset, we combine the information contained in panels (a) and (b) of Figure 1. First, we subtract the log page views of a player the day before a match from the log median page views of the same player in the year before the same match. Second, we subtract from this value the equivalent value for their opponent. As such, our Wiki Relative Buzz Factor measures whether the interest in a player's Wikipedia profile page was atypical the day before a match, and how much it was atypical relative to their opponent in the match. Precisely, for player-match observation i we calculate:

$$\text{WikiBuzz}_i = \ln(w_i/\tilde{w}_i) - \ln(w_{-i}/\tilde{w}_{-i}) \,, \tag{1}$$

where w_i is the previous day's page views for the player, $\tilde{w}_i$ is the median daily page views over the past year, and $-i$ denotes the player's opponent in the match. This measure is plotted in panel (c) of Figure 1 only for the winning player observations in the dataset, being in this case on average negative and significantly predicting a player's defeat (p-value < 0.01). By construction, the variable is mean zero over all winners and losers in the dataset, but we can reject normality with standard tests, due to excess kurtosis of 3.4.

We use the Wikipedia profile page views from the day before the match to construct the buzz factor, instead of the day of the match, because the daily views are in UTC. If we instead used page views from the day of the match, then we could not be confident that the buzz factor was not caused by the outcome of the match, and we could not use it to inform a realistic betting strategy to test market efficiency. In theory, there could still be a very small number of matches in our dataset that began very early in Oceania or Far East Asia before the end of the previous UTC day. We use the match location and timezone data to address this in a robustness check later.

2.2 Detecting mispricing

Let y_i equal one if a player won the match and zero otherwise, where $i = 1, ..., I$. Let p_i be the unobserved beliefs of the bookmaker about the probability of y_i happening beforehand. The bookmaker offers decimal odds o_i on the outcome, meaning that on taking a £1 bet they return o_i to the bettor if the outcome happens and they gain £1 if it does not. Let $z_i = 1/o_i$ be the inverse odds or implied odds-based probability forecast of the bookmaker. For any match, summing z_i over the two players will give a value greater than one, which reflects the bookmaker's expected rate of commission or profit margin κ, also known as the 'overround' or 'vig'. This implies $z_i = p_i + \kappa$. If we denote $e_i = y_i - z_i$, then an efficient bookmaker market requires $E_i[e_i] = -\kappa$. In other words, the the bookmaker is efficient if it makes some average level of commission across matches and outcomes, and no other information can predict e_i, since it will already be priced into the odds.

We consider three potential sources of mispricing and departures from the Efficient Markets Hypothesis in WTA betting markets. First, there is an empirical irregularity in some prediction and betting markets known as the favourite-longshot bias, whereby odds appear to underestimate the chances of the most (least) expected outcomes over the least (most), making bets on favourites more (less) profitable than on longshots (see the summaries by Ottaviani and Sørensen, 2008 and Newall and Cortis, 2021). Most studies of professional sports betting markets have found a longhsot bias, including the seminal study on horse-racing by Ali (1977). Several theoretical contributions have demonstrated the sufficient conditions such that the longshot bias can arise in equilibrium, in terms of preferences, budget constraints and the distribution of beliefs among market participants (e.g., He and Treich, 2017; Manski, 2006; Ottaviani and Sørensen, 2015). In general, high risk aversion can lead to the bias reversing toward the favourite outcome in the market. Besides these predictions from neoclassical theory, a competing set of behavioural explanations has been proposed to explain the bias, which emphasises the misperception of probabilities by bettors (e.g., Snowberg and Wolfers, 2010; Vaughan Williams et al., 2018). Newall and Cortis (2021) suggest from their review of the empirical literature that sports markets with fewer potential outcomes tend to produce a favourite bias (e.g., team sports or tennis), whereas a longshot bias appears in markets with many outcomes (e.g., horse racing or golf). Nevertheless, previous studies of professional tennis have found a longhsot bias (e.g., Abinzano et al., 2016, 2019; Forrest and Mchale, 2007; Lahvička, 2014), though not sufficient to suggest market inefficiency through positive mean returns from consistently betting on match favourites (e.g., Forrest and Mchale, 2007; Štefan Lyócsa and Výrost, 2018).

Second, we consider whether tennis betting markets systematically misprice the outcome of a match according to player rankings. Several studies have demonstrated how the recent performances of tennis players can provide relatively accurate forecasts compared with those implied by bookmaker odds as a benchmark, typically through enhanced Elo (1978) ratings (e.g., Angelini et al., 2021a; Kovalchik and Reid, 2019; Kovalchik, 2020) - we use standard and

more advanced Elo ratings later to provide benchmark probability forecasts of match results.[4] There is some suggestive evidence that bookmakers are more risk averse in tennis matches involving lower ranked players and the longshot bias thus increases in these cases (Abinzano et al., 2016; Lahvička, 2014). The WTA world rankings are ordered from one, for the best player cumulatively over the past year, to having no rank, for a player who has not earned enough points at WTA events over the past year to get one. We consider two measures based on these rankings. First, we consider the raw rank difference between players in a match, $\text{RankDiff}_i = rank_i - rank_{-i}$. Second, we assume that the performance difference between two consecutive players in the rankings is exponentially decreasing as one goes down the ranking list from the top. The difference in ability between the 1st and 2nd ranked players is likely to be more than between the 100th and 101st ranked players, which can be evidenced by how much less often player rankings move at the top compared with the bottom. We construct a ranking distance measure as:

$$\text{RankDist}_i = -\left(\frac{1}{rank_i} - \frac{1}{rank_{-i}}\right), \tag{2}$$

where we impute $1/rank_i = 0$ if a player was unranked at the time of a match. RankDist_i is bounded by –1, when the player considered is ranked first in the world and is playing somebody unranked, and 1, when it is the other way around, thus having the same sign interpretation as RankDiff_i.

Third, we consider our Wikipedia Relative Buzz Factor. To the best of our knowledge, this sort of information has not been used to predict the outcome of tennis matches and the efficiency of their betting markets, or at least this has not been documented before. However, there are parallels with studies using information from social media and player evaluations to predict football outcomes and betting inefficiencies (e.g., Brown et al., 2018; Peeters, 2018).

To detect mispricing and estimate the conditional mean effects on bookmakers' odds implied probability forecast errors, we apply the general Mincer and Zarnowitz (1969) forecast

[4]The Elo ratings are computed using all WTA tennis matches betwene teh beginning of the 2007 season and March 2020.

evaluation framework (see Angelini and De Angelis, 2019, Angelini et al., 2021b, and Elaad et al., 2020, who tested for home bias, the favourite-longshot bias and the weak form efficiency of European football betting markets in much the same way). We estimate the following using least squares:

$$e_i = \alpha + \beta_1 p_i + \beta_2 \text{RankDist}_i + \beta_3 \text{WikiBuzz}_i + \psi_{S(i)} + \phi_{T(i)} + \varepsilon_i \ , \tag{3}$$

where $\{\alpha, \beta_1, \beta_2, \beta_3, \psi_{S(i)}, \phi_{T(i)}\}$ are parameters. We expect a significantly negative estimate of α to capture the bookmaker's profit margin. Positive values of β_1, β_2 or β_3 would respectively suggest a longshot bias, a high-rank bias, and a low-buzz bias in the markets, such that betting on a win by the favourite, the lower ranked player, or the one with greater pre-match relative buzz, could be profitable strategies, and vice versa if these parameters are negative. We also consider fixed effects in Equation (3) for the season (year), $\psi_{S(i)}$, and tournament, $\phi_{T(i)}$, where $S(i)$ and $T(i)$ are indicator functions, to address the potential heterogeneity over these dimensions in bookmaker overrounds or expected profit margins. The remaining heterogeneity is left in the residual term ε_i. We construct standard errors for the estimates of Equation (3) that are robust to clusters at the match and tournament levels. This addresses the heteroskedasticity from including both players in a match in the estimation sample, as well as the possibility that some tournaments are less predictable than others.[5]

The mean of e_i will be significantly negative for any reasonable sized sample of matches. Therefore, a sufficient condition for the betting market to be weak form efficient, according to Equation (3), is given by the null hypothesis: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. If we find that estimates of $\{\beta_1, \beta_2, \beta_3\}$ are significantly positive or negative, then the associated variables provide information that is not fully incorporated in the pre-event prices. In this case, the markets may by inefficient if bettors can use the same information to make sustained positive returns.

[5] As a robustness check, we also considered estimates of Equation (3) using weighted least squares, with elements of the diagonal weighting matrix approximated by $p_i(1-p_i)$, as suggested by Angelini and De Angelis (2019). Although this estimator reduces the influence of more competitive matches, the results that follow are robust to using this instead of ordinary least squares.

2.3 Market inefficiency and a simple betting strategy

To test whether the bookmaker markets are inefficient, we use estimation results of the mispricing model in Equation (3) and the Kelly (1956) criterion. This criterion is the solution to a bettor's maximisation problem on how much of her wealth she should invest in the claim offered by the bookmaker, assuming logarithmic utility and given her beliefs about the outcome of the claim and the odds posted by the bookmaker. Along with simpler strategies, such as "bet one unit when the expected return is positive", the Kelly criterion has been widely used in the literature to evaluate betting market efficiency (e.g., Hvattum and Arntzen, 2010; Peeters, 2018; Ziemba, 2020). We assume that our bettor in this case forms her expectations from estimating Equation (3) using least squares, though without including the season or tournament fixed effects in the model as these are impractical for forecasting. The other variables in Equation (3), and thus estimates of the parameters, are all available to the bettor before a tennis match begins. The bettor's expected probability of winning a bet on event i, a specific player to win a match, is then given by:

$$\widehat{y}_i = \widehat{\alpha} + (1 + \widehat{\beta}_1) p_i + \widehat{\beta}_2 \text{RankDist}_i + \widehat{\beta}_3 \text{WikiBuzz}_i \, . \tag{4}$$

The Kelly criterion gives the share of wealth to invest in the bet, remembering that o_i are the decimal odds offered:

$$z_i = \max\{\widehat{y}_i - \frac{1 - \widehat{y}_i}{o_i - 1} \, , \, 0\} \, . \tag{5}$$

The bettor's return on investment (ROI) over N potential bets, expressed as a percentage of the total amount invested, is given by:

$$\text{ROI} = \frac{\sum_i^N (z_i o_i \mathbb{1}\{y_i = 1\} - z_i \mathbb{1}\{y_i = 0\})}{\sum_i^N z_i} \, . \tag{6}$$

A substantially positive ROI, over a large sample of matches used to generate the estimates in Equation (4), would provide some evidence that tennis match betting markets are weak form inefficient due to some combination of the biases captured by the model. A substantially

positive ROI from out-of-sample forecasting and the application of the betting strategy would be a more powerful result. This would suggest that the relatively straightforward model and betting strategy could be applied profitably in real time. To provide benchmark ROIs, we construct alternative estimates of $\widehat{y}_i$ using the standard player form-based Elo (1978) ratings, with an updating factor (K-factor) of twenty, and using all WTA match results since the beginning of the 2007 season. We also use the more sophisticated W-Elo forecasting model from Angelini et al. (2021a).

3 Results

3.1 Mispricing

Table 1 shows the results of estimating Equation (3), using as the dependent variable the mean pre-match odds offered by the 40-60 bookmakers listed by oddsportal.com. Column (I) only tests for a favourite-longshot bias. We find on average a marginal favourite bias, but this is not statistically significant. Column (II) adds as a regressor the difference in the pre-match WTA rankings of the players, RankDiff_i, which is also not statistically significant. When taken together with the favourite-longshot bias, the null $H_0 : \beta_1 = \beta_2 = 0$ cannot be rejected, and there is no evidence that bookmaker betting markets for WTA tennis matches are mispriced according to the raw difference in ranks and the balance of the odds between players.

In column (III) of Table 1, we replace RankDiff_i with our alternative measure of the rank distance between players, RankDist_i. This measure significantly predicts the average bookmaker odds-implied forecast errors (p-value $= 0.026$). The model estimates suggest that the probability of an unranked player winning against the number one ranked player in the world is 0.054 greater than what bookmaker odds tend to imply. In this specification, there is a small conditional longshot bias, consistent with the previous literature (Abinzano et al., 2016, 2019; Forrest and Mchale, 2007; Lahvička, 2014), though here it is not statistically significant. Taking these two potential sources of mispricing together, the null $H_0 : \beta_1 = \beta_2 = 0$ can be rejected at the 10% level.

In column (IV), we add the third potential source of mispricing to the model in the form of the Wiki Relative Buzz Factor. This measure positively and significantly predicts the average bookmaker odds-implied forecast errors (p-value $= 0.012$). Bookmakers under-predict the likelihood of a win by the player with a relatively larger pre-match increase in Wikipedia profile page views than their opponent. After including this source of mispricing in the model, the estimated rank distance mispricing remains positive and significant, and the longshot bias remains on its own insignificant. We can also reject the sufficient condition for weak form market efficiency, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ at the 5% level. In column (V), we add tournament fixed effects to the regression model and the estimates and test results are practically unchanged.[6]

There could be a very small number of matches in our dataset that began early in Oceania or Far East Asia before the end of the previous UTC day. In these cases, the Wikipedia Relative Buzz Factor could be endogenous and capture the impact of the match outcome, perhaps being correlated with an unexpected result. To address this, column (I) of Table 2 repeats the model estimates from column (IV) of Table 1, and then columns (II)-(IV) show results after cumulatively dropping from the estimation sample matches in time zones from the East, starting with UTC+11&12 (Sydney/Auckland), then UTC+11&12 (Seoul/Tokyo), and finally UTC+7&8 (Singapore/Hong Kong). The influence of the Wiki Relative Buzz Factor and the rejection of the sufficient condition of weak form efficiency are robust to dropping these matches from the estimation sample. After dropping matches from all six of the eastern time zones, the mispricing in odds predicted by the buzz factor is greater. The Wikipedia profile page views less than 24 hours before the start of a match may be less useful in predicting odds mispricing. This would be consistent with the buzz factor being a proxy for crowd judgements on the relative strengths of players' most recent performances within a tournament. To test whether this could alone explain why the buzz factor can predict bookmaker mispricing, in column (V) of Table 2 we re-estimate the model only for matches in the first round of

[6]We checked for misspecification of Equation (3) using Ramsey RESET tests and did not reject the null hypothesis. The data generating process was not better approximated by including squared terms for any of the regressors.

tournaments. The coefficient on the Wiki Relative Buzz Factor remains statistically significant (p-value $= 0.031$) and is larger than when it is estimated over all matches in tournaments. This suggests that the mispricing is not only driven by whatever happened in the previous round of a tournament, which may have generated interest in a player's Wikipedia profile page.

The results from estimating Equation (3), and the tests of mispricing by bookmakers, suggest that there may be inefficiencies in the final result markets of tennis matches. These inefficiencies could be proven by betting on players who are substantially lower ranked then their opponents or who have unusually high interest in their Wikipedia profiles before matches.

3.2 Market inefficiency and the betting strategy

Table 3 shows the results of applying the simple betting strategy described in Section 2.3, by using match outcome probability predictions according to Equation (4) and applying the Kelly criterion. The upper panel of Table 3 applies the strategy in the model estimation sample of matches, for the 2015-20 seasons. The lower panel of Table 3 estimates the model up to the end of the 2018 season, uses this to forecast match outcomes in the 2019 and 2020 season, and then applies the Kelly criterion with these forecasts. Column (I) of Table 3 shows the results of the betting strategy for a hypothetical bettor who could place bets at the average pre-match odds offered by the 40-60 bookmakers sampled for each match. The average overround in these markets in 2015-20 was 5.6%. The in-sample model estimates and Kelly criterion results suggest betting on 554 of the 21,044 considered odds (10,522 matches), with a Return on Investment (ROI) of -5.2%, which is only a small improvement on the average bookmaker overround. When applying the model estimates and strategy out-of-sample, since the start of the 2019 season, a larger proportion of matches are bet on but the ROI is lower at -6.4%. In column (II), motivated by the evidence from Table 2 that the Wiki Relative Buzz factor was a stronger predictor of bookmaker forecast errors once matches in the East were dropped, we estimate the model and apply the betting strategy only for the approximate two-thirds of WTA matches that took place to the west of the time zone containing Singapore and Hong Kong

(UTC+8). In this case we find an in-sample ROI of 2.7% and an out-of-sample ROI of 4.9%, where the latter is generated from 269 bets, or 16% of the matches.

For curiosity, column (III) of Table 3 presents results whereby the model was estimated and predictions were made using average odds, but the best available odds listed on oddsportal.com were used in the Kelly criterion. In this case, the vast majority of matches are bet on and the in-sample and out-of-sample ROIs are respectively 36.1% and 3.1%. However, despite the existence of 'oddschecker' websites being available to the bettor, using the best available odds just before a match begins is not normally realistic, due to the transaction costs and time involved with managing a large number of online accounts. Further, there are restrictions that can prevent a bettor from obtaining the best available odds listed on oddsportal.com, such as the location of a bettor affecting which online sportsbooks they can use. This is evidenced in the lower panel of Table 3, which shows that the average overround according to the best available odds in the 2019 and 2020 WTA seasons was negative, suggesting that theoretical arbitrage opportunities were common if not entirely practical.

As a more realistic test of bookmaker inefficiency, column (IV) of Table 3 presents results from using the Kelly criterion and the odds from only one online sportsbook. We selected Bet365 because it is the highest revenue sportsbook in the world and had odds listed on oddsportal.com for almost every WTA match since 2015. From using the model's predictions and the Bet365 odds, we find in-sample and out-of-sample ROIs of 9.8% and 17.3%, respectively, where the latter value was generated from bets on 12% of the matches since 2019. To check whether these profitable opportunities are driven by the Wiki Relative Buzz Factor, we drop the rank distance measure from the model estimation, with the results in column (V). In this case, fewer matches are bet on according to the Kelly criterion, but the ROIs are increased to 18.1% in-sample and 28.8% out-of-sample. To provide a meaningful benchmark ROI using an alternative probability forecasting model of match results, along with the Kelly criterion, the same samples of matches and the Bet365 odds, column (VI) shows results using the standard Elo (1978) ratings model described in Section 2.3. The ROI

from applying the betting strategy with this alternative set of probability forecasts is -13.3% in-sample and -12.2% out-of-sample. As a further comparison, column (VII) shows betting results using W-Elo, which is a more sophisticated Elo forecasting model of tennis match results that reflects contributions by Kovalchik (2016) and Angelini et al. (2021a). This model gives greater weight to past match wins at prestigious tournaments and takes into account the margins of victory that players achieved.[7] However, the W-Elo model predictions, applied with the Kelly criterion and Bet 365 odds, compared to the standard Elo model in column (VI), generate a higher but still negative in-sample ROI, and a marginally worse out-of-sample ROI.

We check whether the betting returns from using the Wiki Relative Buzz Factor are driven by sub-sets of matches expected to be more or less competitive by bookmakers. We estimate the same model and follow the same betting strategy as in column (V) of Table 3, which yielded an in-sample ROI of 18.1%, except we consider matches in particular odds ranges. We only report in-sample results due to the reduced sample sizes. The results in Table 4 show that applying the model and betting strategy over all sample odds generates a higher ROI than applying it only to the sets of matches that were expected to be relatively competitive or uncompetitive. However, the results also suggest that higher cumulative betting returns may be possible from focusing the model and strategy on intermediate odds, as a higher proportion of bets would then be placed according to the Kelly criterion. In this way, the Wiki Relative Buzz factor tends to be a stronger predictor of bookmaker mispricing when matches are expected to be more competitive, and the players involved are by implication more similar in their ability or form.

In summary, a buzz factor about tennis players, constructed from their Wikipedia profile page views data, provides relevant information that is not being fully incorporated into the match result prices offered by bookmakers. This information can be used to generate sustained and substantial profits when used in a relatively simple betting strategy.

[7] To generate these ratings, we use an R package associated with Angelini et al. (2021a), *welo* (Candila, 2021). When calculating the W-Elo ratings, we restrict the data to only players who played at least 10 WTA matches since the beginning of 2007. The parameters are set to those preferred by Angelini et al. (2021a): player starting points of 1,500, Kovalchik (2016) scale factors, and weights based on the number of games won rather than sets.

4 Conclusion

In this paper, we constructed a measure of relative pre-match buzz about the players in Tennis matches using Wikipedia profile page views data. We found that this Wikipedia Relative Buzz Factor can predict bookmaker forecast errors and the significant mispricing of outcomes, suggesting profitable opportunities for bettors who back a player with relatively greater buzz than their opponent going into a match. Using these results to forecast outcome probabilities and the Kelly criterion to select how much to bet on what matches, we found that tennis result betting markets are inefficient. Prices fail to fully incorporate the information contained in the buzz factor. The returns on investment from applying the model and betting strategy were sustained and substantial, especially when using the odds of Bet365, the world's highest revenue online sportsbook. Two previous studies also found that online information representing the wisdom of crowds can be used to inform profitable betting strategies, though with much smaller rates of return than we have found in tennis markets (Brown and Reade, 2019; Peeters, 2018). However, it is unclear whether correcting these sources of inefficiency would result in greater profits for bookmakers. What we have labelled as mispricing may correlate with unobserved biases and heterogeneity among bettors that bookmakers exploit when setting odds.

There are two natural extensions to this research. The 'wisdom of crowds' might explain why a measure constructed from Wikipedia page views data can predict bookmaker mispricing. While this is an appealing and plausible explanation, we have done nothing here to prove it. This would require complementary data sources that capture explicit predictions about tennis match outcomes or evaluations of the players. The Wikipedia Relative Buzz Factor may only be capturing relative changes in the media interest in tennis players before matches. If that were the case, then our results could perhaps be described more accurately as being driven by the 'wisdom of the media', or by a small number of tennis commentators and pundits who selectively draw attention to some players over others. Second, we can think of no good reason why the betting market inefficiencies found here would be constrained to the top level of

women's professional tennis. It would be interesting for others to check whether these results apply to tennis below the WTA level, men's tennis, or entirely different sports.

References

Abinzano, I., L. Muga, and R. Santamaria. 2016. "Game, set and match: the favourite-long shot bias in tennis betting exchanges." *Applied Economics Letters*, 23(8): 605–608.

Abinzano, I., L. Muga, and R. Santamaria. 2019. "Hidden Power of Trading Activity: The FLB in Tennis Betting Exchanges." *Journal of Sports Economics*, 20(2): 261–285.

Ali, M. M. 1977. "Probability and Utility Estimates for Racetrack Bettors." *Journal of Political Economy*, 85(4): 803–815.

Angelini, G., V. Candila, and L. De Angelis. 2021a. "Weighted Elo rating for tennis match predictions." *European Journal of Operational Research*, Forthcoming.

Angelini, G., and L. De Angelis. 2019. "Efficiency of online football betting markets." *International Journal of Forecasting*, 35(2): 712–721.

Angelini, G., L. De Angelis, and C. Singleton. 2021b. "Informational efficiency and behaviour within in-play prediction markets." *International Journal of Forecasting*, Forthcoming.

Avery, C. N., J. A. Chevalier, and R. J. Zeckhauser. 2016. "The "CAPS" Prediction System and Stock Market Returns." *Review of Finance*, 20(4): 1363–1381.

Brown, A., D. Rambaccussing, J. J. Reade, and G. Rossi. 2018. "Forecasting With Social Media: Evidence From Tweets On Soccer Matches." *Economic Inquiry*, 56(3): 1748–1763.

Brown, A., and J. J. Reade. 2019. "The wisdom of amateur crowds: Evidence from an online community of sports tipsters." *European Journal of Operational Research*, 272(3): 1073–1081.

Candila, V. 2021. *welo: Weighted and Standard Elo Rates*. R package version 0.1.0.

Candila, V., and A. Scognamillo. 2018. "Estimating the Implied Probabilities in the Tennis Betting Market: A New Normalization Procedure." *International Journal of Sport Finance*, 13(3): 225–242.

Chen, H., P. De, Y. J. Hu, and B.-H. Hwang. 2014. "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media." *Review of Financial Studies*, 27(5): 1367–1403.

del Corral, J., and J. Prieto-Rodríguez. 2010. "Are differences in ranks good predictors for Grand Slam tennis matches?" *International Journal of Forecasting*, 26(3): 551–563.

Easton, S., and K. Uylangco. 2010. "Forecasting outcomes in tennis matches using within-match betting markets." *International Journal of Forecasting*, 26(3): 564–575.

Elaad, G., J. J. Reade, and C. Singleton. 2020. "Information, prices and efficiency in an online betting market." *Finance Research Letters*, 35.

Elo, A. E. 1978. *The rating of chessplayers, past and present*. London Batsford.

Fama, E. F. 1965. "The Behavior of Stock-Market Prices." *The Journal of Business*, 38(1): 34–105.

Fama, E. F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance*, 25(2): 383–417.

Forrest, D. 2008. "Soccer Betting in Britain." In *Handbook of Sports and Lottery Markets*. Eds. by D. B. Hausch, and W. T. Ziemba, San Diego Elsevier, 421–446.

Forrest, D., and I. Mchale. 2007. "Anyone for Tennis (Betting)?" *The European Journal of Finance*, 13(8): 751–768.

Galton, F. 1907. "Vox Populi."

He, X.-Z., and N. Treich. 2017. "Prediction market prices under risk aversion and heterogeneous beliefs." *Journal of Mathematical Economics*, 70(C): 105–114.

Hvattum, L. M., and H. Arntzen. 2010. "Using ELO ratings for match result prediction in association football." *International Journal of Forecasting*, 26(3): 460–470.

Kelly, J. L. 1956. "A new interpretation of information rate." *The Bell System Technical Journal*, 35(4): 917–926.

Kovalchik, S. 2020. "Extension of the Elo rating system to margin of victory." *International Journal of Forecasting*, 36(4): 1329–1341.

Kovalchik, S. A. 2016. "Searching for the GOAT of tennis win prediction." *Journal of Quantitative Analysis in Sports*, 12(3): 127–138.

Kovalchik, S., and M. Reid. 2019. "A calibration method with dynamic updates for within-match forecasting of wins in tennis." *International Journal of Forecasting*, 35(2): 756–766.

Kraaijeveld, O., and J. De Smedt. 2020. "The predictive power of public Twitter sentiment for forecasting cryptocurrency prices." *Journal of International Financial Markets, Institutions and Money*, 65(C): .

Lahvička, J. 2014. "What causes the favourite-longshot bias? Further evidence from tennis." *Applied Economics Letters*, 21(2): 90–92.

Štefan Lyócsa, and T. Výrost. 2018. "To bet or not to bet: a reality check for tennis betting market efficiency." *Applied Economics*, 50(20): 2251–2272.

Manski, C. 2006. "Interpreting the predictions of prediction markets." *Economics Letters*, 91(3): 425–429.

McHale, I., and A. Morton. 2011. "A Bradley-Terry type model for forecasting tennis match results." *International Journal of Forecasting*, 27(2): 619–630.

Mincer, J., and V. Zarnowitz. 1969. "The evaluation of economic forecasts." In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. NBER, 1–46.

Newall, P. W. S., and D. Cortis. 2021. "Are Sports Bettors Biased toward Longshots, Favorites, or Both? A Literature Review." *Risks*, 9(1): , p. 22.

Ottaviani, M., and P. N. Sørensen. 2008. "The Favorite-Longshot Bias: An Overview of the Main Explanations." In *Handbook of Sports and Lottery Markets*. Eds. by D. B. Hausch, and W. T. Ziemba, San Diego Elsevier, 83–101.

Ottaviani, M., and P. N. Sørensen. 2015. "Price Reaction to Information with Heterogeneous Beliefs and Wealth Effects: Underreaction, Momentum, and Reversal." *American Economic Review*, 105(1): 1–34.

Peeters, T. 2018. "Testing the wisdom of crowds in the field: Transfermarkt valuations and international soccer results." *International Journal of Forecasting*, 34(1): 17–29.

Scheibehenne, B., and A. Broder. 2007. "Predicting wimbledon 2005 tennis results by mere player name recognition." *International Journal of Forecasting*, 23(3): 415–426.

Snowberg, E., and J. Wolfers. 2010. "Explaining the Favorite-Long Shot Bias: Is it Risk-Love or Misperceptions?" *Journal of Political Economy*, 118(4): 723–746.

Sprenger, T. O., A. Tumasjan, P. G. Sandner, and I. M. Welpe. 2014. "Tweets and Trades: the Information Content of Stock Microblogs." *European Financial Management*, 20(5): 926–957.

Surowiecki, J. 2004. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Brown Little.

Vaughan Williams, L. 1999. "Information Efficiency in Betting Markets: A Survey." *Bulletin of Economic Research*, 51(1): 1–30.

Vaughan Williams, L., M. Sung, P. A. F. Fraser-Mackenzie, J. Peirson, and J. E. V. Johnson. 2018. "Towards an Understanding of the Origins of the Favourite–Longshot Bias: Evidence from Online Poker Markets, a Real-money Natural Laboratory." *Economica*, 85(338): 360–382.

Ziemba, W. T. 2020. "Parimutuel betting markets: racetracks and lotteries revisited." SRC Discussion Paper 103, Systemic Risk Centre, London School of Economics.

TABLE 1: Model estimates and tests of betting market mispricing for WTA match results, 2015-2020

	(I)	(II)	(III)	(IV)	(V)
Odds-implied probability	-0.009	-0.040	0.013	0.036	0.036
	(0.020)	(0.034)	(0.022)	(0.024)	(0.024)
WTA rank diff. (player-opponent)		-0.010			
		(0.008)			
WTA rank distance to opponent			0.054**	0.049**	0.049**
			(0.024)	(0.027)	(0.024)
Wiki Relative Buzz Factor				0.009**	0.009**
				(0.003)	(0.003)
Constant	-0.025**	-0.009	-0.037***	-0.049***	-0.047***
	(0.011)	(0.018)	(0.012)	(0.013)	(0.013)
Year/season fixed effects	Yes	Yes	Yes	Yes	No
Tournament fixed effects	No	No	No	No	Yes
F-test: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$		0.436	0.076	0.016	0.016
N of players	21,044	20,992	21,044	21,044	21,044

Notes.- ***,**,* indicate significance from zero at 1%, 5% and 10% levels, respectively, two-sided tests. Standard errors in parentheses were estimated robust to both match and tournament level clusters.
Column (I): linear regression estimates of Equation (3), where the dependent variable is the forecast error implied by average bookmaker odds (oddsportal.com) – test of favourite-longshot bias
Column (II): adds the pre-match raw WTA rank difference to the model in (I)
Column (III): uses the alternative differences in ranks measure described in the text – the coefficient effect should be interpreted as an unranked player against the number one ranked in the world, relative to two hypothetically equally ranked players
Column (IV): adds the Wiki Relative Buzz Factor – preferred results
Column (V): adds tournament fixed effects to the model in (IV)

TABLE 2: Model estimates and tests of betting market mispricing for WTA match results, 2015-2020: preferred model and dropping time zones, and 1st round matches only

	(I)	(II)	(III)	(IV)	(V)
Odds-implied probability	0.036	0.034	0.037	0.050*	0.065**
	(0.024)	(0.026)	(0.026)	(0.029)	(0.032)
WTA rank distance to opponent	0.049**	0.055*	0.059**	0.048	0.059
	(0.024)	(0.029)	(0.029)	(0.032)	(0.054)
Wiki Relative Buzz Factor	0.009**	0.008**	0.009**	0.012***	0.012**
	(0.003)	(0.004)	(0.004)	(0.004)	(0.006)
Constant	-0.049***	-0.048***	-0.050***	-0.056***	-0.065***
	(0.013)	(0.014)	(0.014)	(0.015)	(0.006)
Drop UTC+11&12	No	Yes	Yes	Yes	No
Drop UTC+9&10	No	No	Yes	Yes	No
Drop UTC+7&8	No	No	No	Yes	No
Year/season fixed effects	Yes	Yes	Yes	Yes	Yes
F-test: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$	0.016	0.021	0.011	0.008	0.063
N of players	21,044	19,032	17,872	14,940	9,596

Notes.- ***,**,* indicate significance from zero at 1%, 5% and 10% levels, respectively, two-sided tests. Standard errors in parentheses were estimated robust to both match and tournament level clusters.
Column (I): repeats the preferred model estimates from column (IV) of Table 1
Column (II)-(IV): each column drops matches from an additional two time zones, starting with UTC+11&12 (Sydney/Auckland) and finally in column (IV) dropping UTC+7&8 (Singapore/Hong Kong)
Column (V): estimates the preferred model from column (I) but only using matches from the first round of tournaments

TABLE 3: Betting strategy results for WTA match results, within the model estimation sample and out-of-sample

	Average odds		Best	Bet365			
	All (I)	w/out East (II)	(III)	All (IV)	w/out rank (V)	Elo (VI)	W-Elo (VII)
In-sample, 2015-2020:							
N odds ($2 \times N$matches)	21,044	14,940	21,042	20,984	20,984	19,685	12,198
Number of bets placed	554	824	9,556	935	742	7,308	5,701
Mean overround (%)	5.57	5.56	0.12	6.64	6.64	6.67	6.57
Return on Investment (%)	-5.20	2.70	36.10	9.81	18.13	-13.26	-11.45
Out-of-sample, 2019-2020:							
N odds ($2 \times N$matches)	5,190	3,416	5,188	5,156	5,156	4,796	4,362
Number of bets placed	221	269	2,350	312	276	1,778	2,058
Mean overround (%)	5.33	5.34	-0.23	6.46	6.46	6.48	6.49
Return on Investment (%)	-6.37	4.87	3.05	17.26	28.82	-12.24	-12.38

Notes.- "In-sample" uses the full sample of matches from column (V) of Table 1, the model predictions and applies the Kelly criterion. "Out-of-sample" instead estimates the model up to the end of the 2018 season and then uses it to predict match outcomes and apply the Kelly criterion for the 2019 & 2020 seasons. Average odds are always used to estimate the models and generate forecasts, but the odds used in the Kelly criterion are varied.
Column (I): uses the reported average or pre-match available odds from oddsportal.com
Column (II): uses the reported average or pre-match available odds from oddsportal.com but only for matches taking place in time zones to the west of UTC+7 (Singapore)
Column (III): uses the reported best available pre-match odds from oddsportal.com
Column (IV): uses pre-match odds from Bet365
Column (V): uses Bet365 odds but with a version of the preferred model estimated without the rank distance variable
Column (VI): uses Bet365 odds but with the standard Elo predicted probability forecast of the match outcome
Column (VII): uses Bet365 odds but with the W-Elo predicted probability forecast of the match outcome as per Angelini et al. (2021a).

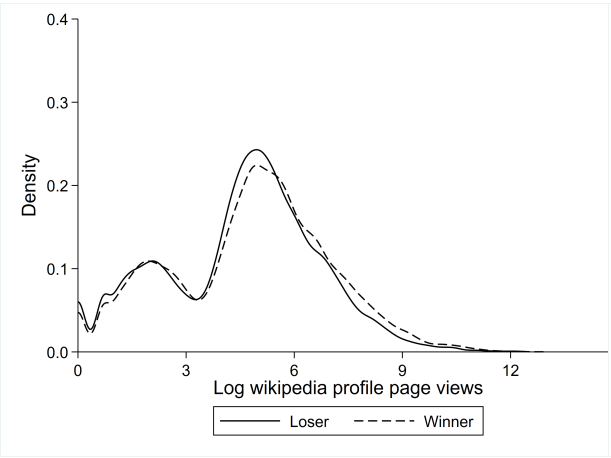
TABLE 4: Betting strategy results for WTA match results, within the model estimation sample: selecting sample odds based on match competitiveness

	Bet365 odds			
	(I)	(II)	(III)	(IV)
In-sample, 2015-2020:				
N odds ($2 \times N$matches)	3,224	14,200	17,770	6,800
Number of bets placed	63	143	1,091	1,121
Mean overround (%)	5.76	6.33	6.80	7.30
Return on Investment (%)	6.03	16.35	11.83	8.88

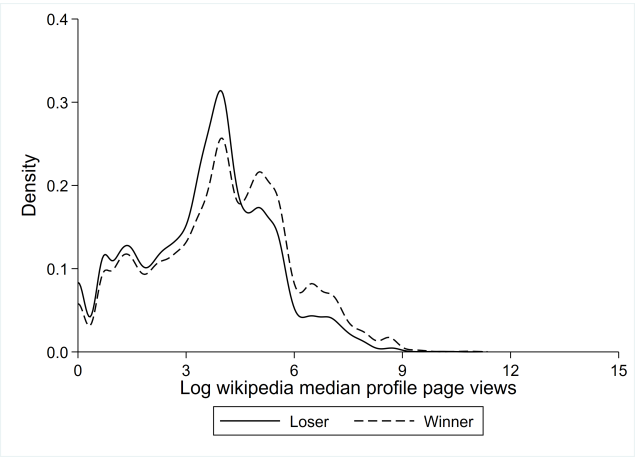
Notes.- Betting strategy results equivalent to Column (V) of Table 3, varying the sample of match odds used in the model and considered for bets by column. Average odds are always used to estimate the models and generate forecasts, but Bet365 odds are used in the Kelly criterion.
Column (I): uses only odds in the sample which imply a match win probability of $p \in (0, 0.2) \cup (0.8, 1)$
Column (II): uses only odds in the sample which imply a match win probability of $p \in (0, 0.4) \cup (0.6, 1)$
Column (III): uses only odds in the sample which imply a match win probability of $p \in [0.2, 0.8]$
Column (IV): uses only odds in the sample which imply a match win probability of $p \in [0.4, 0.6]$

FIGURE 1: Wikipedia page views of tennis players before WTA matches in 2015-2020

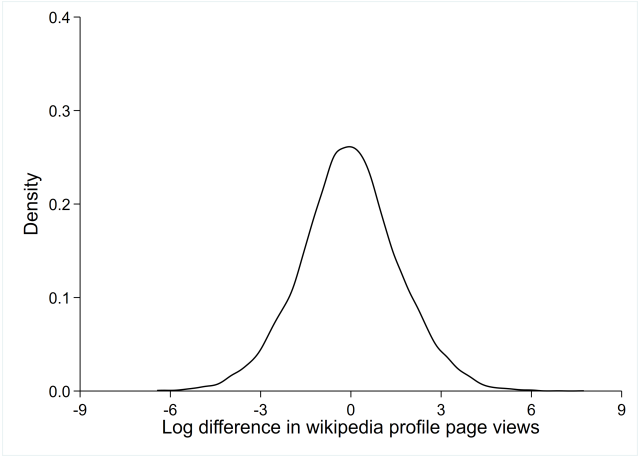
(a) The day before the match



(b) Median in the year before the match



(c) Relative buzz factor: Log difference between the winner's page views yesterday and their median daily views in the year before, relative to the loser



Notes: author calculations using Wikipedia Foundation pageview data for the English language profiles of WTA tennis players, collected daily (Coordinated Universal Time (UTC)) Wikipedia page views of their English language profiles using the Pageview Application Programming Interface (API). The density is estimated with a Gaussian kernel and bandwidth of 0.2.