University of Zurich[UZH]

Department of Business Administration

UZH Business Working Paper Series

Working Paper No. 384

Sonic Thunder vs Brian the Snail

Fast-sounding racehorse names and prediction accuracy in betting exchange markets

Oliver Merz, Raphael Flepp and Egon Franck

October 2019

---

http://www.business.uzh.ch/forschung/wps.html

UZH Business Working Paper Series

Contact Details

Oliver Merz

oliver.merz@business.uzh.ch

Raphael Flepp

raphael.flepp@business.uzh.ch

Egon Franck

egon.franck@business.uzh.ch

University of Zurich

Department of Business Administration

Affolternstrasse 56, CH-8050 Zurich, Switzerland

Sonic Thunder vs Brian the Snail
Fast-sounding racehorse names and prediction accuracy in betting exchange markets

Oliver Merz, Raphael Flepp, and Egon Franck*

October 2019

*Corresponding author: Oliver Merz, Affolternstrasse 56, CH-8050 Zurich, Switzerland. oliver.merz@business.uzh.ch. Coauthors: Dr. Raphael Flepp, University of Zurich, Affolternstrasse 56, 8050 Zurich, Switzerland, and Prof. Dr. Egon Franck, University of Zurich, Affolternstrasse 56, 8050 Zurich, Switzerland.

Abstract
This paper examines the influence of objectively irrelevant information on prediction accuracy in horse-racing betting exchange markets. In horse racing, the name of a horse does not depend on the horse's performance and is thus uninformative. We investigate the impact of fast-sounding horse names on prediction market price accuracy and betting returns. Using over 3 million horse bets, we find evidence that the winning probabilities of bets on horses with fast-sounding names are overstated, which impairs the prediction accuracy of such bets. This finding implies that the prices in betting exchange markets are not efficient, as prices become distorted by incorporating the misleading information from a horse's fast-sounding name. This bias translates into significantly lower betting returns for horses classified as fast-sounding compared to the returns of all other horses.

Keywords
Market efficiency · Sports forecasting · Prediction markets · Betting industry · Horse Racing

JEL Classification
D40 · G40 · C53 · L83

1. Introduction

Imagine betting on a horse in a race without properly knowing the past performance or the rankings of the horses involved in the race. You could choose to bet your money on a horse called *Sonic Thunder* or on a horse called *Brian the Snail*. On which horse would you bet your money?

In prediction markets, individuals trade on the outcomes of future events (Brown, Reade & Vaughan Williams, 2019). Thus, as prediction markets are designed to allocate resources and aggregate information, market prices forecast future events (Berg, Nelson & Rietz, 2008). As the participants in the prediction markets profit from accurate predictions, they have strong incentives to acquire useful information, and thus, prediction market prices are expected to accurately forecast the underlying outcome (Brown, Reade & Vaughan Williams, 2019). Many studies have analyzed the accuracy of prediction market prices. For example, Wolfers and Zitzewitz (2004), Berg et al. (2008) and Rothschild (2015) demonstrate that prediction markets are among the most accurate sources for political predictions. Vaughan Williams and Reade (2016) find that prediction markets produce more accurate forecasts than opinion polls and expert opinions in the context of political races, while Spann and Skiera (2009) show that prediction markets are more accurate than tipsters' forecasts in the context of sports. Furthermore, Franck, Verbeek and Nüesch (2010) show that in betting (prediction) markets, betting exchanges provide more accurate predictions than bookmakers.

Similar to financial markets, prediction markets are efficient if the market prices reflect all historical information and the prices are the best forecasts for the outcome probabilities of a match (Angelini & de Angelis, 2019). However, some more recent studies have shown deviations from efficient prices because the prices do not incorporate all available information. For example, Brown, Rambaccussing, Reade and Rossi (2017) find that the aggregate tone of Twitter tweets provides additional information that is not incorporated into

the betting (prediction) market prices. Specifically, they find that an overall positive tone in Tweets indicates that a football team is approximately 3.4% more likely to win the game than implied by the contemporaneous betting prices. Furthermore, Brown and Reade (2017) find that the aggregation of tips from sports tipsters predicts sports outcomes beyond the betting prices. Specifically, they show that if a majority of tipsters predict that a team or individual will win, then the team or the individual is 1.6% more likely to win the game than implied by the betting odds. Taken together, these studies find evidence that there is relevant information available that is not fully included in the prediction market prices and that consideration of this additional information could improve the prediction accuracy of the underlying outcomes. While these examples illustrate cases where not all relevant information is fully reflected by the prediction market prices, there might be circumstances where irrelevant information is incorporated into the prices, harming the prediction accuracy.

In this paper, we investigate whether irrelevant information impairs prediction accuracy in a betting (prediction) market. In particular, we analyze whether the name of a horse has an impact on price accuracy and thus market efficiency. The name of a horse such as *Sonic Thunder* or *Brian the Snail* is uninformative because the regulations concerning the naming of a horse do not consider past success. Rather, the name of a horse usually cannot be changed after the horse has participated in a race. Thus, the name of the horse has no signaling effect regarding its performance.[1]

Nevertheless, Tversky and Kahneman (1974) explain that to simplify decision making, people often rely on mental shortcuts when faced with complicated choices. Thus, a casual bettor might simply make her/his decision based on the most obvious attribute of a horse, its name. If the bettor associates the horse's name with speed and thus expects that the horse will run fast, then the bettor might already be inclined to bet on this horse before making an

[1] Indeed, our empirical results show that the correlation of horses with fast-sounding names and actual wins is basically zero (-0.002).

extensive assessment of all publicly available factors such as the offered odds, past placements or the physical appearance of the horse. Slovic, Peters, Finucane and MacGregor (2007) state that affective responses occur rapidly and automatically and that people quickly associate feelings with stimulus images or words such as "treasure" or "hate". Therefore, the bettor might already have subconsciously experienced a positive affect towards a fast-sounding horse before evaluating the horse's chance of winning.

If there is a large enough share of casual inexperienced bettors who bet on a horse that has a positive affect on them due to its fast-sounding name, then we expect to find deviations from efficient prices and price accuracy impairments because the disproportionate demand for such bets is likely to lead to higher (lower) prices (odds). Thus, we expect less accurate prices for bets on fast-sounding horses. If those bettors wrongly value the name of a horse as a quality signal, they pay a premium for bets on fast-sounding horses. Consequently, we expect lower returns for bets on fast-sounding horses compared to bets on other horses.

We use data from over 400,000 horse races between 2008 and 2018 with more than 3 million horse bets. Data are obtained from the betting exchange Betfair, where bettors trade bets against each other in a continuous double auction. Following previous research, e.g., Forrest and Simmons (2008) and Franck, Verbeek and Nüesch (2011), we use logit regressions with the outcome of a bet as the dependent variable (equaling 1 if the bet is won and 0 if lost) and as the explanatory variables, we use the probabilities implied in the odds[2] and a binary variable indicating whether a horse is classified as fast-sounding. If the odds (prices) are efficient, all relevant information should be reflected in them, and no additional variables should have predictive power regarding the outcome of an event.

We find that having a fast-sounding horse name has predictive power with regard to the race outcome beyond the probabilities implied in the odds. In particular, our results show

[2] As in previous research, e.g. Forrest and Simmons (2008) and Franck et al. (2011), we calculate probabilities of an event occurring by the reciprocal of the odds.

that the winning probabilities of bets on horses with fast-sounding names are overstated, implying that the prices in betting exchange markets are not completely efficient as prices become distorted by incorporating misleading or false information. Furthermore, we find significantly lower returns for horses classified as fast-sounding compared to other horses. A simple trading strategy of betting against all horses classified as fast-sounding yields a return of approximately 2.9% before commission but a negative return of -1.6% after commission[3]. This finding could be bracketed under the "limits of arbitrage" argument of Gromb and Vayanos (2010) because the mispricing is not large enough to overcome the transaction costs; thus, potentially misleading or false information is not fully eliminated from prices. Nevertheless, this strategy generates significantly larger profits than a random betting strategy where zero returns are achieved before commission and a negative return of -4.7% after the commission is considered. Despite wagering real money, a substantial share of the betting community seems to be systematically biased in preferring bets on fast-sounding horses over bets on other horses. This finding supports the view that irrelevant factors can impair price accuracy and market efficiency to some degree. Overall, this paper contributes to the general discussion of the accuracy of prediction markets and extends the previous literature by examining the effect of objectively irrelevant factors on prediction market prices. Although prediction markets are highly effective at allocating resources and aggregating information, their forecasting accuracy is impaired if large crowds with biased valuations are present.

The remainder of this paper is structured as follows. In section 2, we describe the data, the process of classifying fast-sounding horses and the empirical methodology used. In section 3, we present our results and a trading strategy. Section 4 concludes the paper.

[3] We applied the default market base rate of 5% for UK and Ireland from Betfair.

2. Methods

2.1 Data

We collected betting data on horse races from Paddy Power Betfair, one of the largest online wagering operators in the world (Paddy Power Betfair PLC., 2018)[4]. Betfair operates markets across various sports, as well as in politics and economics (Paddy Power Betfair PLC., 2018). The betting exchange mechanism mirrors a standard limit order book in financial markets, where traders can submit limit and market orders in a continuous double auction (Flepp, Nüesch, and Franck, 2017). Thus, bettors can choose whether they want to place a bet that an event will occur (backing a bet) or bet against the event occurring (laying a bet). If two parties with opposing opinions agree on a price, their bets are matched and a transaction takes place (Franck et al., 2010). Furthermore, Betfair offers the option to back or to lay a horse at an ex-ante unknown starting price (SP). At the time the race begins, Betfair determines the market clearing starting price from the aggregate volumes of back and lay bets.[5] For each race, we obtained data on the date, time and location of the race, the names of the participating horses, the winner of the race, the weighted average matched price (WAP) and the starting price (SP) at the end of the prerace period. The WAP and the SP are both denoted in decimal odds.

We collected data on 443,850 horse races held in the UK, Ireland, USA, South Africa and Australia from March 2008 until May 2018. With an average number of approximately 9.2 horses per race, we observe a total of 4,066,445 horses. We exclude all races where more than one horse won simultaneously and all races where we do not observe all odds for the participating horses because the values are either missing or erroneous[6]. Thus, the final

[4] The data from Betfair are freely available on http://www.Betfairpromo.com/Betfairsp/prices

[5] The starting prices are calculated by Betfair to match the betting volumes from the back and lay sides. The idea of the starting price is to generate a price that matches the largest share of the betting volume.

[6] In approximately 900 races, there were multiple winners. Our results are insensitive to the inclusion of those races. A value for the WAP or SP is deemed erroneous if the price is below 1.01 or above 1000, which corresponds to the odds range given by Betfair. Our results do not change if we only exclude the missing values for WAP or SP instead of the whole race.

samples used in our analysis consist of 344,749 races and 3,193,458 horses using the WAP and 422,816 races and 3,903,604 horses for the analysis using the SP. Summary statistics for the betting odds of WAP and SP are depicted in Table 1.

Table 1: Summary statistics for betting odds

	Observations	Number of Races	Mean	SD	Min	Max
WAP	3,193,458	344,749	38.41	86.20	1.01	1000
SP	3,903,604	422,816	47.32	106.87	1.02	1000

Notes: WAP is the weighted average of the traded prices prior to the race based on their traded volume. SP is the Betfair starting price that is calculated by Betfair based on the volumes from backing and laying customers.

2.2 Fast-sounding name categorization

The categorization of horse names into fast-sounding and not fast-sounding is of critical importance. To come up with an objective list of terms that are associated with speed, we used four different sources. In particular, we used two commonly employed and very popular thesauri, the Cambridge and Oxford dictionaries, a word association API called *twinwords* and www.horses-names.com, an independent website that provides suggestions for horse names, to evaluate whether a horse's name is associated with speed[7]. Using the Cambridge and Oxford dictionaries, we looked up synonyms for the words "fast" and "speed" and then classified a horse as fast-sounding if one of those terms was contained in the horse's name[8]. Using the *twinwords* API, we captured horses whose name included a word associated with the terms "speed" or "fast". Additionally, we used the name suggestions for fast horses from www.horses-names.com to classify fast-sounding horses. While the dictionaries and

[7] The data for the different sources is freely accessible and was retrieved from the following URLs.
Cambridge Dictionary: https://dictionary.cambridge.org/topics/moving-quickly-and-slowly/fast-and-rapid/
Oxford Dictionary: https://en.oxforddictionaries.com/thesaurus/fast and https://en.oxforddictionaries.com/thesaurus/speed
Twinwords API: https://www.twinword.com/api/word-associations.php
Horse names website: http://www.horses-names.com/fast-horse-names.php

[8] Unfortunately, the Cambridge Dictionary does not provide a synonym search for speed. However, the term "speed" comes up as a synonym for "fast". The Oxford Dictionary allows for a more thorough approach as it distinguishes between adverbs and adjectives for "fast" as well as between nouns and verbs for "speed". Further, the Oxford Dictionary groups synonyms depending on the context. We only include terms that relate to speed of movement. We did not include words declared by Oxford to be vulgar, informal, rare, literary or archaic as these are likely to be misinterpreted.

word association API capture synonyms for "speed" and "fast", this website also provides words that are commonly associated with speed, e.g., Rocket or Comet. A comprehensive list of the terms determined to indicate a fast-sounding name can be found in the Appendix in Tables 9-12.

A horse is classified as fast-sounding if its name includes any of the terms on the list, e.g., "Speed Dragon" would be classified as fast-sounding because of the word "Speed"[9]. Table 2 shows 40 selected fast-sounding horses to further illustrate which names are classified as fast-sounding.

Table 2: Fast-sounding horse name examples

Sonic Power	Speed Dragon	Zippy Lad	Lightening Vault
Powerful Jet	Orbit Express	Swift Chap	Blazing Tempo
Brave Falcon	Rush Now	Top Magic	Dixie Flyer
Esprit De Bullet	Strike Fast	Hustle Hard	Diamond Rush
Crown Me Fast	Hot Seat	Top Gear	Bright Bullet
Quick Art	Rush Of Blood	Top Boy	Meteoric Moments
One Wild Guy	Sonic Thunder	Grand Gallop	Zippy Speed
Run for Roses	Saratoga Wildcat	Quick Beers	Sudden Rush
Flyingwithoutwings	Fast On	Dazzlem Quick	You Drive I Fly
Irish Rocket	Hot Sauce	Mighty Flying Thomas	Fullshot

Notes: To illustrate the types of names that are classified as fast-sounding, 40 names have been selected.

[9] For the classification of horse names, we only consider words starting with a capital letter, e.g., we only search for "Speed" but not "speed". With this approach, we can capture almost all relevant names because our data provide the names with a capital letter between spaces. At the same time, this approach avoids cases where preceding letters change the meaning of a word, e.g. the term "top" would also capture "stop".

The number of horses classified as fast-sounding resulting from each source is shown in Table 3. Panel A shows the individual contributions of the four sources to the share of horses categorized as fast-sounding when using the WAP, and Panel B analogically shows the corresponding contributions to the share of horses categorized as fast-sounding when using the SP. The underlying reason for the categorization approach used in this paper is twofold. First, it is important to obtain a comprehensive list of terms indicating speed in order to capture a large share of horses with fast-sounding names and thus enable comparisons between the two groups. As shown in Table 3, the individual sources capture relatively small shares of horses classified as fast-sounding. However, if we combine all sources additively, we are able to establish a more comprehensive subsample. As some classifications overlap, we have a final subsample of fast-sounding horse names of approximately 4.3% of observations for the WAP variable and approximately 4.6% for the SP variable.

Table 3: Sample composition

Panel A: Share of horses classified as fast-sounding using WAP					
Classified as fast-sounding	Cambridge Dictionary	Oxford Dictionary	Twinwords	Horses-names.com	Combined
0	3,133,678 (98.13%)	3,120,071 (97.70%)	3,165,247 (99.12%)	3,150,886 (98.67%)	3,055,314 (95.67%)
1	59,780 (1.87%)	73,387 (2.30%)	28,211 (0.88%)	42,572 (1.33%)	138,144 (4.33%)
Panel B: Share of horses classified as fast-sounding using SP					
Classified as fast-sounding	Cambridge Dictionary	Oxford Dictionary	Twinwords	Horses-names.com	Combined
0	3,829,343 (98.10%)	3,810,480 (97.61%)	3,867,983 (99.09%)	3,846,086 (98.53%)	3,726,182 (95.45%)
1	74,261 (1.90%)	93,124 (2.39%)	35,621 (0.91%)	57,518 (1.47%)	177,422 (4.55%)

Notes: This table shows the share of horses classified as fast-sounding across different sources. Panel A shows the distributions of horses classified as fast-sounding for the WAP and Panel B shows the distributions of horses classified as fast-sounding for the SP.

Second, the use of four independent and complete sources helps to mitigate the potential subjectivity of the name categorization process. Using the whole list of words associated with speed inevitably leads to some errors in categorization, as some words might not be distinctively related to movement speed and could be interpreted differently depending on the context. However, to ensure an objective classification procedure, we refrained from excluding individual words that are potentially subject to misinterpretation and always used the complete list of words associated with "fast" and "speed" according to the four sources.

2.3 Statistical Methods

Following previous research, e.g., Forrest and Mchale (2007), Forrest and Simmons (2008) and Flepp, Nüesch and Franck (2016), we use the reciprocal of the decimal odds to calculate the market's forecasting probability of a certain bet to win. We calculated the implied winning probabilities $impliedprob_{i,wap} = \frac{1}{WAP}$ and $impliedprob_{i,SP} = \frac{1}{SP}$ for each horse. A favorite horse that is more likely to win has a higher implied winning probability as it trades at lower odds. For example, a horse with an SP of 1.25 is expected to win in $\frac{1}{1.25} =$ 80% of the cases.

We follow Franck et al. (2011) and Forrest and Simmons (2008) by examining whether our indicator variable for fast-sounding horse names has explanatory power beyond the implied probabilities with regard to the actual outcomes. We test a binary model with the actual outcome of a bet (1 for a winning bet; 0 for a losing bet) as the dependent variable; the implied probability and our indicator variable (fast-sounding) are explanatory variables. Specifically, we estimate our multivariate logit model as follows:

$$prob(bet\ i\ wins) = f(impliedprob_i, fast - sounding_i) \quad (1)$$

where *impliedprob* is the probability that is implied by the WAP or SP and *fast-sounding* is a binary variable indicating whether a horse's name is classified as fast-sounding or not.

Under the null hypothesis, we assume that betting exchange markets are efficient and thus the prices fully reflect all available information, including the names of the horses. In other words, we assume the prices to be the best outcome forecasts of the underlying events. Thus, the coefficient of the fast-sounding variable should be zero. If our indicator variable has explanatory power in addition to the implied probabilities, then the odds are not efficient and bets on horses with fast-sounding names are not equally profitable, on average, as bets on other horses. We expect a negative sign for the fast-sounding variable if a large enough share of bettors with a biased preference towards fast-sounding names demand bets on those horses. As we include multiple observations of the same race (bets on the participating horses), the independence assumption between those observations is violated. To account for this, we compute clustered heteroscedasticity-robust standard errors at the race level.

In an alternative approach, we use t-tests and nonparametric Wilcoxon rank-sum tests to check whether betting returns significantly differ among bets on horses that have a fast-sounding name and horses without a fast-sounding name. We calculate returns of one-unit bets using the formula $return_i = \frac{stake \times odds_i - stake}{stake}$ where $odds_i$ represent either the WAP or the SP. Because the returns are not normally distributed, i.e., for approximately 89% of the observations, the return is -1 whenever a bet is lost, we additionally conduct Wilcoxon rank-sum tests. We determine the returns by conducting one-unit bets on horses to win a race. If markets are rational and efficient, there should be no systematic difference in returns between any subgroups of bets.

3. Results

3.1 Price accuracy of fast-sounding horse bets

Table 4 depicts summary statistics and correlation coefficients for the variables *impliedprob* and *fast-sounding*. The correlation coefficients indicate that there is almost no relationship between fast-sounding horses and well-performing horses. If anything, fast-

sounding horses seem to be slightly less likely to win because the correlation coefficient between Win_i and $Fast\text{-}sounding_i$ is negative. The correlations between fast-sounding horses and the probabilities implied in the odds are slightly positive for the WAP and slightly negative for the SP. Overall, the correlations are close to zero, indicating a weak relationship.

Table 4: Summary statistics and correlation coefficients

Panel A: Sample using WAP							
Variable	Mean	SD	Min	Max	1	2	3
1 Win_i *(0/1)*	0.1078	0.3102	0.0000	1.0000	1.0000		
2 $impliedprob_{i,WAP}$	0.1185	0.1270	0.0010	0.9901	0.3681	1.0000	
3 $Fast\text{-}sounding_i$	0.0433	0.2034	0.0000	1.0000	-0.0018	0.0046	1.0000
Panel B: Sample using SP							
Variable	Mean	SD	Min	Max	1	2	3
1 Win_i *(0/1)*	0.1083	0.3110	0.0000	1.0000	1.0000		
2 $impliedprob_{i,SP}$	0.1088	0.1205	0.0010	0.9804	0.3856	1.0000	
3 $Fast\text{-}sounding_i$	0.0455	0.2083	0.0000	1.0000	-0.0019	-0.0013	1.0000

Notes: Panel A shows the summary statistics and correlation coefficients using the WAP to calculate implied probabilities and Panel B for the SP. Win_i represents the actual outcome of bet *i* (0/1), *impliedprob* the probability odds, and *fast-sounding* is an indicator variable for fast-sounding horse names.

The results of the logit regression are depicted in table 5. The results are shown in the form of marginal effects measured at a point where the continuous *impliedprob* variable is set to its mean and the binary *fast-sounding* variable is set to zero. Panel A in table 5 shows the results using the WAP to calculate implied probabilities, and Panel B shows the results using the SP to calculate the implied probabilities. The results for the two estimations for the *impliedprob* variable stemming from WAP and SP are consistent. The sign for *impliedprob* is positive and significant at the 1% significance level for both estimations. More importantly, the sign of the variable *fast-sounding* is negative and significant at the 1% significance level for the WAP implied probabilities and negative and significant at the 5% significance level for the SP implied probabilities. Thus, the information about fast-sounding horses is not correctly reflected in the market odds (prices), and the null hypothesis of market efficiency is rejected. The variable *fast-sounding* has a significant impact in predicting a win and has a

negative sign, implying that the implied probabilities for fast-sounding horses are too high. This result suggests that horses with fast-sounding names are overvalued by bettors.

Table 5: Results of logit regressions

Panel A: Using winning probabilities implied by WAP

		Win (0/1)
impliedprob$_{WAP}$		0.509***
		(0.001)
fast-sounding		-0.005***
		(0.001)
Number of observations	3,193,458	
Number of clusters	344,749	
Pseudo R^2	0.145	
Log pseudolikelihood	-933625.89	

Panel B: Using winning probabilities implied by SP

		Win (0/1)
impliedprob$_{SP}$		0.553***
		(0.001)
fast-sounding		-0.002**
		(0.001)
Number of observations	3,903,604	
Number of clusters	422,816	
Pseudo R^2	0.158	
Log pseudolikelyhood	-1126406.1	

Notes: The dependent variable *Win* is binary equaling 1 if a horse has won the race or 0 if the horse did not win the race. Marginal effects of the variables *impliedprob* and *fast-sounding* are depicted. Heteroscedasticity-robust and clustered standard errors at the race level are reported in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

To test the robustness of our results, we conducted several variations of our main specification. First, as an alternative to the logit model that assumes a logistic distribution, we

ran the regressions using a probit model and standard OLS and obtained virtually the same results (not shown for brevity). Second, we randomly selected a bet on one horse per race as an alternative to using clustered standard errors and obtained very similar results. Finally, following Forrest & Simmons (2008), we adjusted the implied probabilities of the horses such that they sum to one for each particular race. Again, our results remain virtually the same.

3.2 Comparison of betting returns

Using two-sided t-tests and nonparametric Wilcoxon rank-sum tests, we examine how the returns differ on average for horses classified as fast-sounding compared to horses not classified as fast-sounding. Table 6 shows that on average, the returns for fast-sounding horses are significantly lower than the returns of the other horses. Panel A shows the results using the WAP to calculate returns, and Panel B shows the results using the SP to calculate returns. Independent of the odds used to calculate the returns, we observe significantly lower returns for bets on horses that are classified as fast-sounding. The returns for bets on fast-sounding horses are approximately 4.1 percentage points lower when using the WAP, and the difference is significant at the 1% significance level. Using the SP leads to similar results, a 3.6-percentage-point lower return for bets on fast-sounding horses, and the difference is significant at the 5% significance level. Nonparametric Wilcoxon rank-sum tests confirm the results, with the differences being significant at the 1% significance level for both WAP and SP. Overall, the results of the t-tests and Wilcoxon rank-sum tests suggest that the returns of bets on fast-sounding horses are systematically lower than the returns of bets on horses without a fast-sounding name.

Table 6: Betting returns comparisons for bets on fast-sounding horses and bets on other horses

Panel A: Returns using the WAP

		t-test			Wilcoxon rank-sum test		
	N	Mean	SE	t	rank sum	expected	z
Fast-sounding	138,144	-0.1610	0.0139		$2.199 \cdot 10^{11}$	$2.206 \cdot 10^{11}$	
Other	3,055,314	-0.1206	0.0030		$4.879 \cdot 10^{12}$	$4.879 \cdot 10^{12}$	
Δ	3,199,790	-0.0405	0.0146	-2.780***			-3.536***

Panel B: Returns using the SP

		t-test			Wilcoxon rank-sum test		
	N	Mean	SE	t	rank sum	expected	Z
Fast-sounding	177,422	-0.0289	0.0155		$3.454 \cdot 10^{11}$	$3.463 \cdot 10^{11}$	
Other	3,726,182	0.0067	0.0038		$7.274 \cdot 10^{12}$	$7.273 \cdot 10^{12}$	
Δ	3,903,604	-0.0356	0.0179	-1.988**			-3.656***

Notes: The table displays the results of simple two-sided t-tests and Wilcoxon rank-sum tests based on the two groups of horses whose names are classified as fast-sounding and on other horses. Panel A shows the returns using the WAP, and Panel B shows the returns using the SP.

Because we observe lower returns for fast-sounding horses, we can derive a simple trading strategy to exploit this finding. Due to the nature of betting exchange markets with no intermediary, the losses of the losing bettors equal the gains of the winning bettors. Thus, our findings imply positive returns for betting against all horses classified as fast-sounding[10]. Table 7 shows the returns of backing fast-sounding horses and laying fast-sounding horses before and after commission costs are considered. As lay bets work differently compared to back bets, the potential gains and losses differ. For a back bet, the maximum loss a bettor risks is its stake, e.g., 1 unit in our case. For lay bets, a bettor potentially risks much more than its stake. The potential loss of a lay bet is called the liability and can be calculated using the formula: $liability = stake \times odds - stake$. Table 7 illustrates that for back bets, a bettor never loses more than 1 unit, while for lay bets, the largest loss equals 739 units. In our

[10] A bettor could simply choose to bet against all horses classified as fast-sounding at the Betfair starting price. Although the WAP shows even stronger results, it could not be used to form a trading strategy as the WAP is the volume weighted average of the odds traded during the pre-play period that is unknown to a bettor ex-ante.

sample using the same stakes, returns for back bets are characterized by many small losses and few large gains, while the returns of lay bets are characterized by few large losses and many small gains.

Table 7: Returns for back and lay bets on fast-sounding horses

Panel A: Returns before commission					
	N	Mean	SD	Min	Max
Back return	177,422	-0.0289	6.5348	-1	739
Lay return	177,422	0.0289	6.5348	-739	1
Panel B: Returns after commission					
	N	Mean	SD	Min	Max
Back return	177,422	-0.0722	6.2147	-1	702.05
Lay return	177,422	-0.0158	6.5281	-739	0.95

Notes: The table displays the returns of back and lay bets with a stake equaling 1. Panel A shows returns before commission and panel B shows returns after a commission rate of 5% has been deducted.

Theoretically, a return of approximately 2.9% could be achieved by simply betting against all fast-sounding horses. However, the returns turn to -1.6% after incorporating the Betfair commission rate of 5%[11]. Thus, the mispricing in the odds for fast-sounding horses is not large enough to overcompensate the commission costs imposed by Betfair. Although our proposed strategy of laying fast-soundings horses is not economically viable, the betting returns are significantly higher compared to the returns of a random betting strategy.[12]

4. Conclusion

The purpose of this paper is to examine whether objectively irrelevant factors influence price accuracy in a prediction market environment where prices are formed by bettors trading with each other. Specifically, we analyze whether a horse name that indicates

[11] 5% is the default market base rate for UK and Ireland at Betfair.
[12] For this strategy, we randomly selected 177,422 bets from our entire sample. After commission, this strategy generated a lay return of approximately -4.7%.

speed affects bettor behavior and consequently the forecasting accuracy of prices on the betting exchange.

We find that the winning probabilities of bets on horses with fast-sounding names are overstated, which impairs the prediction accuracy of such bets. Thus, the prices in the betting exchange market are not efficient because they incorporate misleading information from a horse's fast-sounding name. Furthermore, we find that returns on horses with fast sounding names are systematically lower compared to the returns of all other horses. This result suggests that one should avoid jumping on the bandwagon when many other bettors are tempted to base their investment decisions on irrelevant factors and instead be aware of a potential mispricing of such bets.

A potential limitation of this paper is that the classification process of determining fast-sounding horses is not completely flawless. Although objective, our classification procedure unavoidably evokes two types of classification errors. First, we might wrongly classify some horses as fast-sounding because some synonyms do not solely relate to movement speed. Second, we probably fail to correctly classify some fast-sounding horses because our sources do not capture all synonyms and related words to "fast" or "speed". As such, our results are likely to be biased towards zero, and a more stringent approach might yield more compelling results. For example, a full list of horse names could be handed out to a wide range of experts who then assess each name individually.

This paper contributes to the general discussion of prediction market accuracy and extends it by analyzing the impact of irrelevant factors on price accuracy and market efficiency. While prediction markets are great outlets through which to allocate resources and aggregate information, the presence of large crowds relying on irrelevant information harms their accuracy to some degree.

References

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

Berg, J. E., Nelson, F. D., & Rietz, T. A. (2008). Prediction market accuracy in the long run. *International Journal of Forecasting*, *24*(2), 285-300.

Brown, A., Rambaccussing, D., Reade, J. J., & Rossi, G. (2018). Forecasting with social media: evidence from tweets on soccer matches. *Economic Inquiry*, *56*(3), 1748-1763.

Brown, A., & Reade, J. J. (2017). Combining Prediction Markets and Forecasting Contests. *Available at SSRN 2944675*.

Brown, A., Reade, J. J., & Williams, L. V. (2019). When are prediction market prices most informative?. *International Journal of Forecasting*, *35*(1), 420-428.

Flepp, R., Nüesch, S., & Franck, E. (2016). Does bettor sentiment affect bookmaker pricing? *Journal of Sports Economics*, 17(1), 3-11.

Flepp, R., Nüesch, S., & Franck, E. (2017). The liquidity advantage of the quote-driven market: Evidence from the betting industry. *The Quarterly Review of Economics and Finance*, 64, 306-317.

Forrest, D., & McHale, I. (2007). Anyone for tennis (betting)?. *The European Journal of Finance*, 13(8), 751-768.

Forrest, D., & Simmons, R. (2008). Sentiment in the betting market on Spanish football. *Applied Economics*, 40(1), 119-126.

Franck, E., Verbeek, E., & Nüesch, S. (2010). Prediction accuracy of different market structures—bookmakers versus a betting exchange. *International Journal of Forecasting*, 26(3), 448-459.

Franck, E., Verbeek, E., & Nüesch, S. (2011). Sentimental preferences and the organizational regime of betting markets. *Southern Economic Journal*, 78(2), 502-518.

Gromb, D., & Vayanos, D. (2010). Limits of arbitrage. *Annu. Rev. Financ. Econ*., 2(1), 251-275.

Paddy Power Betfair PLC. (2018). *Annual report & accounts 2017*. Retrivied from https://www.paddypowerBetfair.com/~/media/Files/P/Paddy-Power-Betfair/documents/ppb-r-n-a-2017-180327.pdf

Rothschild, D. (2015). Combining forecasts for elections: Accurate, relevant, and timely. *International Journal of Forecasting*, *31*(3), 952-964.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European journal of operational research*, 177(3), 1333-1352.

Spann, M., & Skiera, B. (2009). Sports forecasting: a comparison of the forecast accuracy of prediction markets, betting odds and tipsters. *Journal of Forecasting*, *28*(1), 55-72.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.

Williams, L. V., & Reade, J. J. (2016). Prediction markets, social media and information efficiency. *Kyklos*, *69*(3), 518-556.

Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of economic perspectives*, *18*(2), 107-126.

Appendix

Table 9: *Cambridge Dictionary* Synonyms for "fast"

a mile a minute	helter-skelter	quick-fire
apace	high-speed	quickly
as fast as your legs would carry you	hot	quickness
as if it is going out of style	hotfoot	rapid
at a rate of knots	hustle	rapid-fire
at full pelt	in the twinkling of an eye	rate
at full speed	Jack Robinson	say
at full tilt	lick	shot
at full tilt	lickety-split	smartly
before you can say Jack Robinson	lightning	souped-up
blistering	like a shot	spanking
breakneck	like a streak of lightning	speed
brisk	like lightning	spread like wildfire
chop-chop	meteoric	streak
crash	mile	style
express	nimble	superfast
fast	nimbleness	supersonic
fleet	nimbly	swift
full	nippy	swiftly
full steam ahead	pdq	thick
gallop	pell-mell	thick and fast
galloping	poky	tilt
go like hot cakes	posthaste	top
have a heavy foot	precipitous	twinkling
headlong	precipitously	whoosh
heavy	prompt	wildfire
hell	promptly	zippy
hell for leather	quick	

Notes: List of synonyms for "fast" from the Cambridge Dictionary, retrieved from: https://dictionary.cambridge.org/topics/moving-quickly-and-slowly/fast-and-rapid/

Table 10: *Oxford Dictionary* Synonyms

Panel A: Synonyms for "fast"		
accelerated	high-speed	pell-mell
at full speed	hurried	post-haste
at full tilt	hurriedly	quick
at speed	in a flash	quickly
at the speed of light	in a hurry	rapid
blistering	in a trice	rapidly
breakneck	in a wink	smart
brisk	in haste	speedily
briskly	in in time	speedy
energetically	in no time at all	sporty
expeditious	in the blink of an eye	sprightly
expeditiously	like a flash	swift
express	like a shot	swiftly
fast	like an arrow from a bow	turbo
fast-moving	lively	unhesitating
fleet-footed	meteoric	whirlwind
flying	nimble	with all haste
hastily	on the double	with dispatch
hasty	pell-mell	without delay

Panel B: Synonyms for "speed"		
acceleration	haste	scutter
alacrity	hasten	sharpness
blast	hurriedness	shoot
bolt	hurry	spank along
bowl along	hurry	speed
briskness	hurtle	speediness
career	immediacy	sprint
celerity	momentum	stampede
charge	pace	streak
dart	precipitateness	sweep
dash	promptness	swiftness
dispatch	quickness	swoop
expedition	race	tempo
expeditiousness	rapidity	uzz
fastness	rate	velocity
flash	rattle along	whirl
fly	run	whizz
gallop	rush	whoosh
go hell for leather	scramble	wing
go like lightning	scud	zoom
hare	scurry	

Notes: List of synonyms for "fast" (both adjectives and adverbs) from Oxford Dictionary, retrieved from: https://en.oxforddictionaries.com/thesaurus/fast
Only included terms regarding movement speed, i.e., first section of adjectives and adverbs. Terms belonging to informal, British informal, North American informal, literary or rare are not included. List of synonyms for "speed" (both nouns and verbs) from Oxford Dictionary, retrieved from: https://en.oxforddictionaries.com/thesaurus/speed
Terms belonging to informal, British informal, Scottish informal, North American informal, North American vulgar slang, literary, archaic or rare are not included.

Table 11: *Twinword API* words associated with "fast" and "speed"

Panel A: Words related to "fast"		
abrupt	impetuous	rushed
agility	outrun	scramble
dash	overhasty	speed
disconcerted	overrun	speedily
dodge	promptly	speedy
haste	quick	sudden
hastily	quickly	suddenly
hurried	rapid	swift
hurriedly	rapidly	swiftly
hurry	rush	zoom
Panel B: Words related to "speed"		
accelerate	haste	race
acceleration	hasten	rapidity
agility	hie	rush
airspeed	hurriedly	speedy
celerity	hurry	stronghold
dash	pace	swift
decelerate	quick	swiftness
expedite	quicken	tempo
fast	quickly	urgently
fastness	quickness	velocity

Notes: Notes: List of words associated with "fast" from the Twinword API, retrieved from: https://www.twinword.com/api/word-associations.php

Table 12: Fast horse-name suggestions from horse-names website

Apache	Bentley	Blustery
Bullet	Buzz	Comet
Cougar	Falcon	Faster
Flash	Ghost rider	Harley
Jet	Jump	Jumping
Miles	Mustang	Pony express
Quick	Quicky	Racer
Rapid	Rapide	Rocket
Sonic	Speedy	Taz
Tornado	Traveler	Wildfire
Voyager	Wild	Velocity

Notes: List of fast-sounding name suggestions from an independent website, retrieved from: http://www.horses-names.com/fast-horse-names.php