The Temporal Evolution of Mispricing in Prediction Markets

Valerio Restocchi[a,b], Frank McGroarty[a], Enrico Gerding[b]

[a]*Southampton Business School, University of Southampton, SO17 1BJ, Southampton, UK*
[b]*Electronics and Computer Science, University of Southampton, SO17 1BJ, Southampton, UK*

Abstract

We analyze mispricing in prediction markets, a powerful forecasting tool that harnesses the wisdom of the crowd. We show that prediction market prices exhibit mispricing, and we quantify its temporal evolution. Our results suggest that level of the FLB, averaged over the entire time period, decreases with market duration, but this changes when considering only the last trading days. In that case, we find FLB to be positively correlated with duration. We argue that this type of temporal dynamics of mispricing we observe is consistent with herding behavior.

Keywords: Mispricing, Prediction markets, Asset Pricing, Favorite-longshot bias
JEL Classifications: D4; G1

1. Introduction

In this paper we analyze mispricing in prediction markets. Prediction markets are effective tools that gauge the *wisdom of the crowd*, thus greatly improving prediction accuracy on a wide range of events [1]. In fact, although prediction markets are most famous for election forecasts, often outperforming polls and experts [2], they have been recently employed by prominent companies such as Google, Microsoft, Intel and many others to improve predictions of a number of key variables, e.g., revenues, sales volume of specific products, company share price, etc. [3, 4, 5]. Also, because they possess a definite end-point at which the outcome of an event is observed, prediction markets represent an ideal test bed to study decision making under uncertainty and investor behavior in financial markets [6].

In our analysis, we focus on the favorite-longshot bias (FLB), an empirical regularity whereby bets on likely (unlikely) outcomes are underpriced (overpriced). The FLB represent an important topic of study, which has attracted researchers for decades [7, 8, 9], mainly because, unlikely other price anomalies, it has not disappeared in time, but it is still exhibited by betting market prices [10]. Previous work has shown that the FLB also exists in prediction markets [11, 4], and provided analysis of the correlation between this type of mispricing and time left to expiration [12]. However, to date, there is no comprehensive analysis of how mispricing evolves over time in prediction markets depending on their duration.

In this paper, we fill this gap by showing that political prediction markets exhibit the FLB in most cases, and we characterize its temporal evolution. Our analysis complements and adds to the one by Page and Clemen [12]. In their paper, they find that mispricing becomes significant only when the time left to expiration is sufficiently long. We differentiate our analysis by providing a complete temporal profile of mispricing (i.e, evaluating the FLB at each trading day) depending on market duration (i.e., the number of days between the start and the end of a market), which allows us to find different dynamics of mispricing over time. Specifically, in contrast to what has been observed by Page and Clemen [12], we find that, during the last days of trading, the FLB is significantly positive and also positively correlated with duration (i.e, the longer the market, the greater the mispricing). Also, our results suggest that, on average, markets are the more efficient the longer they last. That is, the level of the FLB averaged over the entire period is lower the longer the duration.

The remainder of this paper is organized as follows. In Section 2 we outline the data and methods we use to perform our analysis, and in Section 3 we show and describe our findings. In Section 4 we discuss our results and argue that the FLB in the last days is caused by herding.

2. Data and Method

To perform our analysis, we use end-of-day prices from 3363 markets, from October 2014 to November 2016, and constitutes all markets traded on PredicIt[1], a platform that allows betting on the outcomes of political events, during this period.. Each market represents a possible outcome on a political event, and it is possible to buy or sell a contract at a price $0 < \pi < 1$. Such a contract pays 1 dollar if the selected outcome occurs, and 0 dollars otherwise. For example, a user who buys a "Clinton will lead" contract on "Who will lead in Trump vs. Clinton polling on September 14?", will pay π dollars to the seller, who, in

---

*Corresponding author
Email addresses:* `vr2g14@soton.ac.uk` (Valerio Restocchi),
`f.j.mcgroarty@soton.ac.uk` (Frank McGroarty), `eg@ecs.soton.ac.uk`
(Enrico Gerding)

[1]www.predictit.org

turn, at the end of the market (in this case, 14 September 2016), will pay the buyer 1 dollar if Clinton was leading the polls, and nothing otherwise.

By definition, the FLB is observed if prices on the favorite (longshot) are lower (higher) than the corresponding true probability of the outcome to occur. However, it is fair to assume that it is impossible to know the *fair price* of a single contract (i.e., the true probability of the associated outcome to occur). To address this issue, we analyze all markets by comparing the price on a given day with the final outcome. By aggregating these results, we can see the discrepancy between the realized frequencies of positive outcomes and corresponding prices.

The FLB exists if, for low-probability outcomes, prices are systematically higher than the true probability of the associated outcome to occur, and lower otherwise. Formally, prices exhibit the FLB if:

$$\frac{\pi_A}{\pi_B} < \frac{p_A}{p_B} \iff p_A > p_B \tag{1}$$

where π_A and π_B represent the price to buy a contract on outcomes A and B, and p_A and p_B represent the true probabilities of such outcomes to occur. In our analysis, since we cannot know a priori the true probability of an event, p_i is the observed frequency of positive outcomes for contracts with prices $\pi = i$, with $0 < i < 1$. For a sufficiently large sample, p_i tends to the average true probability associated with price π_i, thus it is possible to represent the FLB with the following measure:

$$\Phi = \Phi_- + \Phi_+ \tag{2}$$

where

$$\Phi_- = \sum_{i=0}^{i=0.49} \pi_i - p_i \tag{3}$$

$$\Phi_+ = \sum_{i=0.51}^{i=1} p_i - \pi_i \tag{4}$$

where Φ_+ and Φ_- represent the cumulative difference between prices and true probabilities, for high and low probability events respectively. To reduce noise in the estimation of the FLB, we allow increments of i to vary according to the number of observation we possess at a given point in time (i.e., a given number of days before the end of the market). Such increments are always between 0.02 and 0.1.

3. Empirical Analysis

In this section we examine the temporal dynamics of the FLB. We start by showing that there exists an FLB in political prediction markets when we average over the entire time period. We then examine this in more detail by grouping the markets depending on their duration. Following this, we examine prices at different points in time (i.e., different number of days to expiration), which enables analysis of the temporal evolution of the FLB (see Fig. 1). In these analyses, we consider all markets with at least one day of trading activity in our data set, resulting in a total of 3363 markets.
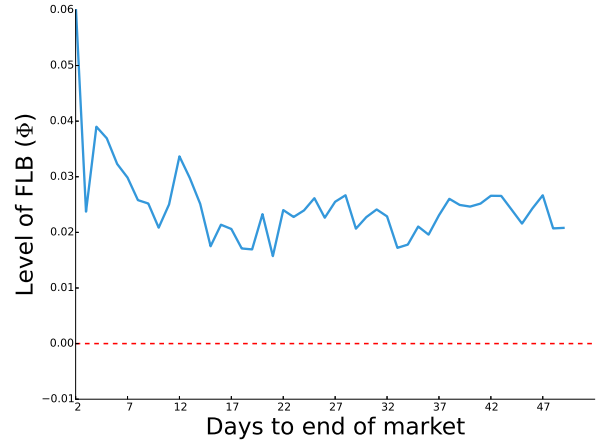

Figure 1: The level of the FLB depending on the number of trading days left to the end of the market, computed considering all markets.

Specifically, we examine prices depending on the number of days to the end of the market, which enables analysis on the temporal evolution of the FLB (see Fig. 1). We find that the average level of the FLB throughout the last 50 days of trading is small but positive, at $\Phi = 0.025 \pm 0.007$. Hence, we perform a t-test to check whether the time series of Φ is significantly different than random zero. These results, shown in Table 1, suggest that there exists a positive FLB in political prediction markets, and that, on average, its level grows significantly during the last 15 days of trading.

To gain more insights, we also group markets depending on their duration and analyze the FLB separately for each group. Results are shown in Table 1 and Fig. 2. We find that prices exhibit the FLB with a high level of confidence for all market groups except those longer than 50 days, for which a low score in the t-test (0.224) and a high p-value (0.06) do not allow us to reject the null hypothesis that mispricing is just noise with a 5% significance. However, we find that, if we measure the FLB within the 24 hours before expiration[2] (denoted with Φ_{last}), this market group shows the highest level of the FLB among all the groups considered. In fact, our results suggest that Φ_{last} increases with the duration of the market, growing from $\Phi_{last} = 0.058$ for markets that last no more than 10 days, to $\Phi_{last} = 0.151$ for markets longer than 50 days. From Table 1, it is also possible to see that long markets display, on average, lower levels of the FLB, and thus they exhibit little mispricing. Specifically, except from very short markets, our results suggest a negative correlation between the average mispricing and that of the last days of trading.

These observations are consistent with the finding that, in prediction markets, volumes grow exponentially towards the

---

[2]To perform our analysis, we use end-of-day prices. Since each market has a definite and different end point during the last day, to measure the FLB in the last trading day we use the last available end-of-day price, which is the close price of the day before the event occurs. Consequently, the FLB measured in this way is the mispricing that exists 0-24 hours before the end of the market

2

Duration (τ, in days)	Markets	Average Φ	St.Dev. Φ	Φ_{last}	t-test
$\tau > 0$	3363	0.025	0.007	0.061	15.92*
$0 < \tau \leq 10$	1784	0.038	0.009	0.058	12.30*
$10 < \tau \leq 25$	600	0.068	0.015	0.089	13.87*
$25 < \tau \leq 50$	435	0.039	0.042	0.136	4.51*
$\tau > 50$	544	0.008	0.030	0.151	1.92**

* = significant at the 0.01%.

** = not significant.

Table 1: Summary statistics for the level of FLB Φ, average value of Φ the day before the end of the market, and t-test statistic for the temporal evolution of Φ.

end of the market. In the data we analyze, short markets are usually based on events of lesser importance, that relate to main events for which it is possible to trade for months. For example, markets such as *"Will Democrats lead by 2.9% or less, or tie or trail Republicans, in generic congressional polling on November 7?"* are open only for a few days and exhibit low liquidity, whereas those on more significsant events such as *"Will Hillary Clinton win the 2016 Democratic presidential nomination?"* can last for a considerably longer time (in this case, two years) and attract a significant amount of bets. Given this, it is possible that the liquidity of longer-term markets is sufficient to reduce or eliminate the FLB. Also, towards the end of the market, important events are extensively covered by both traditional and social media, and it is not rare that for such events, prediction markets prices are also reported. This may trigger herding behavior, and attract people who are not rational and attach a higher probability of realization to the most probable outcome. In this scenario, rational and informed traders would not be able to push the price back to its fundamental value, due to time and liquidity constraints [13], while this is possible during the rest of the market, when liquidity is lower and time is not a constraint to move the price.

4. Conclusions

We analyzed political prediction markets and showed that their prices exhibit a regular mispricing pattern, known as the favorite-longshot bias (FLB). To perform a more detailed analysis of the mispricing, we group them by duration, which allows us to find that the FLB is both related to the time left to expiration and the market duration. In particular, we observe two important trends. First, we find that, on average, the level of the FLB, when taken over the entire period, is higher the shorter the market duration. Second, we find that this dynamic changes during the last 24 hours of trading, when the FLB becomes larger the longer is the market duration, i.e., there is positive correlation between the two quantities. We argue that these observations are consistent with herding behavior and the fact that market efficiency increases the higher the trading volumes. Indeed, on the one hand, prediction markets that allow trading on the outcomes of important events typically have a longer duration, and, on average, important events attract more money (i.e., a higher daily volume). Consequently, the extent of the mispricing in long markets is expected to be lower, as markets which exhibit high volumes attract more informed bettors, who use private information to improve price efficiency [14, 15]. On the other hand, important events receive far more attention than unimportant ones, and gain increasingly more media exposure the closer they get to the end. Moreover, in the recent past, the media use prediction markets to help their forecasts of the probability of specific events to occur. This attracts more bettors to prediction markets, potentially causing herding behavior in the last 24 hours of trading, making it difficult for informed bettors to push the price back to its fundamental value before the end of the market, due to shortage of time. This explanation is in agreement with our observation that, during the last trading day, the FLB is positively correlated to market duration, which is in stark contrast to the findings of Page and Clemen [12] that FLB decreases overtime. Since Page and Clemen use an older data set in their analysis, and consider markets from 10-15 years before ours, we argue that one potential explanation of these conflicting results may be that, nowadays, media make more conspicuous use of prediction markets to support their forecasts, thereby attracting more late bettors which engenders herding. We believe this is an important and interesting issue that requires further investigation in future work, to better understand the mechanisms that generate mispricing in prediction markets, with the ultimate goal of improving predictive power.

[1] J. Berg, F. Nelson, T. Rietz, Prediction market accuracy in the long run, International Journal of Forecasting 24 (2008).

[2] J. Wolfers, E. Zitzewitz, Prediction Markets in Theory and Practice, 2006.

[3] C. Plott, K. Chen, Information aggregation mechanism: concept, design, and implementation for a sales forecasting problem, 2002.

[4] B. Cowgill, J. Wolfers, E. Zitzewitz, Using prediction markets to track information flows evidence from google, in: S. Das, M. Ostrovsky, D. Pennock, B. Szymanski (Eds.), Auctions, Market Mechanisms and Their Applications., Springer, Berlin, lecture no edition, 2009.

[5] D. E. O'Leary, Prediction Markets as a Forecasting Tool, in: K. Lawrence (Ed.), Advances in Business and Management Forecasting, volume 8, Emerald Group Publishing Ltd, London, 2011, pp. 169–184.

[6] K. Croxson, J. J. Reade, Information and Efficiency: Goal Arrival in Soccer Betting, The Economic Journal 124 (2014) 62–91.

[7] R. Griffith, Odds adjustements by american horse-race bettors, The American Journal of Psychology (1949) 9–13.

[8] M. Ottaviani, P. N. Sørensen, Late Informed Betting and the Favorite-Longshot Bias (2003).

[9] E. Snowberg, J. Wolfers, Explaining the FavoriteLong Shot Bias: Is it Risk-Love or Misperceptions?, Journal of Political Economy 118 (2010) 723–746.

[10] R. D. Mclean, J. Pontiff, Does Academic Research Destroy Stock Return Predictability?, The Journal of Finance 71 (2016) 5–32.

[11] P. W. Rhode, K. S. Strumpf, Historical Presidential Betting Markets, Journal of Economic Perspectives 18 (2012) 127142.

[12] L. Page, R. T. Clemen, Do prediction markets produce well-calibrated probability forecast?, The Economic Journal 123 (2012) 491–513.
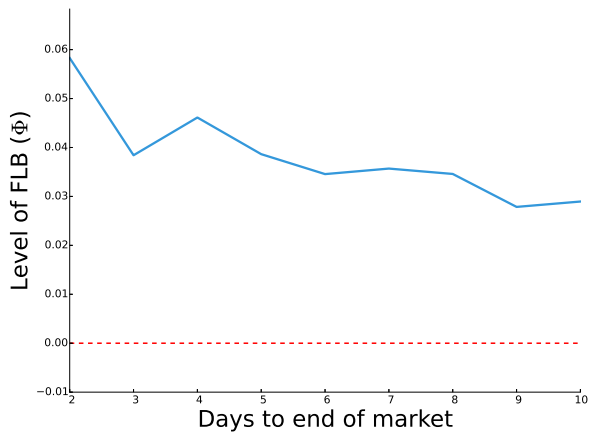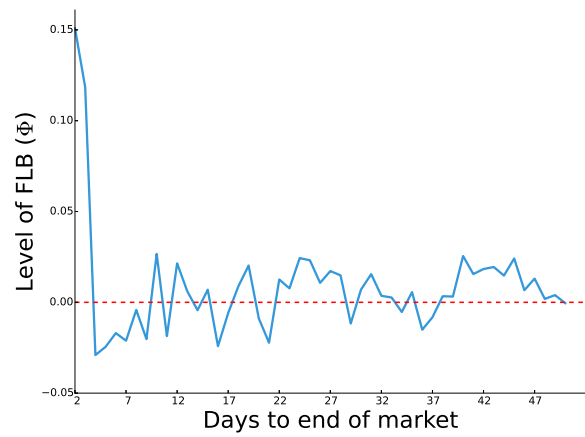
(a) Duration less or equal than 10 days.
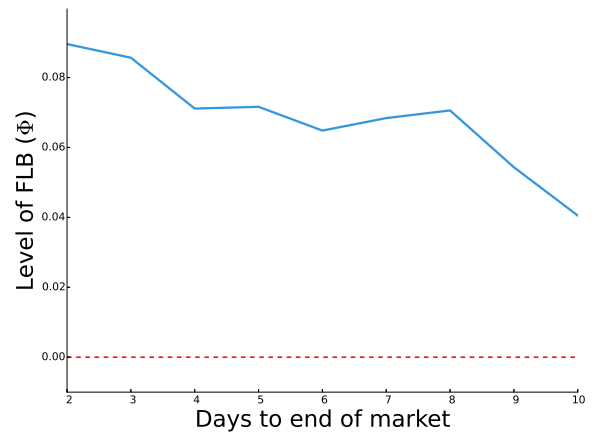
(b) Duration between 10 and 25 days.
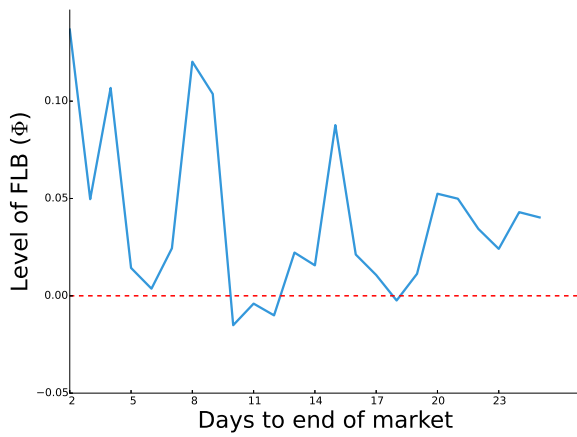
(c) Duration between 25 and 50 days.

(d) Duration longer than 50 days.

Figure 2: Level of the FLB for different market durations.

[13] T. Lux, M. Marchesi, Scaling and criticality in a stochastic multi-agent model of a Financial market, Nature 397 (1999) 498–500.

[14] S. J. Grossman, J. E. Stiglitz, On the impossibility of informationally efficient markets, The American Economic Review 70 (1980) 393–408.

[15] A. R. Admati, P. Pfleiderer, A theory of intraday patterns: volume and price variability, Review of Financial Studies 1 (1988) 3–40.