

It Takes All Sorts: A Heterogeneous Agent Explanation for Prediction Market Mispricing[☆]

Valerio Restocchi^{a,b}, Frank McGroarty^a, Enrico Gerding^b, Johnnie E. V. Johnson^{a,*}

^a*Southampton Business School, University of Southampton, SO17 1BJ, Southampton, UK*

^b*Electronics and Computer Science, University of Southampton, SO17 1BJ, Southampton, UK*

Abstract

Pricing anomalies threaten the value of prediction markets as a means of harnessing the 'wisdom of the crowd' to make accurate forecasts. The most persistent and puzzling pricing anomaly associated with price-implied prediction probabilities is the favourite-longshot bias (FLB). We demonstrate that existing models of the FLB fail to capture its full complexity, thereby preventing appropriate adjustments to market forecasts to improve their accuracy. We develop an agent-based model with heterogeneous agents in a fixed-odds market. Our agent-based simulations and comprehensive analysis using market data demonstrate that our model explains real market behaviour, including that of market makers, better than existing theories. Importantly, our results suggest that adequately complex models are necessary to describe complex phenomena such as pricing anomalies. We discuss how our model can be used to better understand the relation between market ecology and mispricing in contexts such as options and prediction markets, consequently enhancing their predictive power.

Keywords: Forecasting; OR in prediction markets; Agent-based modelling; Cognitive bias; OR in sports.

1. Introduction

Financial instruments such as forward and futures contracts have long been used to reveal people's collective expectations about future outcomes by tapping into demand and supply in the market today to set a price for delivery (settlement) of a commodity (asset value) at some specified future date. The forward contract is one of the most ancient financial instruments in existence. The most recent incarnation of the futures contract, so-called 'political futures' related to the US Presidential election outcomes, have been traded on the Iowa Electronic Markets since 1988. The Iowa market and its emulators, such as Intrade and PredictIt, together with event-based betting (e.g. election outcomes, interest or tax rate changes, sports events etc.) offered by bookmakers and betting exchanges, are known collectively as *prediction markets*. These markets are heralded as effective mechanisms for harnessing the *wisdom of the crowd* to make accurate forecasts (e.g., Berg et al. (2008); Arnesen and Bergfjord (2014)).

[☆]We are grateful to two anonymous referees for providing us with very helpful comments. We also thank the participants to WEHIA'16 and seminars held at University of Southampton and UCL. Finally, we want to thank Vincezo Corallo for invaluable discussion on early versions of this paper.

*Corresponding author

Email addresses: v.restocchi@soton.ac.uk (Valerio Restocchi), f.j.mcgroarty@soton.ac.uk (Frank McGroarty), eg@ecs.soton.ac.uk (Enrico Gerding), j.e.johnson@soton.ac.uk (Johnnie E. V. Johnson)

The ability of the forecasts drawn from prediction markets to fully reflect all relevant information is predicated on the Efficient Market Hypothesis (EMH) (Fama, 1970), which states that market prices always incorporate all relevant information. In particular, this would suggest that those who hold information would continue to trade in prediction markets until they believe that their information is fully assimilated into the prevailing market prices. However, pricing anomalies, which result in market prices failing to fully reflect all relevant information, present a potential threat to the forecasting accuracy of prediction markets. If the mechanisms which lead to a particular pricing anomaly are well understood, then it may be possible to identify when and how to adjust final market prices to annul the impact of the anomaly. However, if overly simplistic models of these anomalies are employed, adjustments based on these models may be wholly inadequate, or in some cases harmful, to the accuracy of resulting forecasts.

We argue that in prediction markets, as much as in other fields, it is important that a model describe a phenomenon with the sufficient degree of complexity. Otherwise, it risks missing those meaningful dynamics that are a consequence of complex interactions between the components of the system (e.g., market participants). Models that do not describe these phenomena precisely are unlikely to accurately forecast prices (e.g., fail to spot early signals of herding behaviour) or, in the case of state-contingent claims markets, such as options and prediction markets, to estimate the true probabilities associated to an event.

For more than a century, it was argued that models with only a representative agent could sufficiently describe systems such as financial markets, in which participants are supposed to be rational. However, with the growth of computational power and the introduction of behavioral economics, many criticise this approach. Indeed, although the representative agent approach has the important merit of making models analytically tractable, it has been argued that the representative agent is an unjustified, incorrect assumption that leads to fundamentally wrong conclusions (Kirman, 1992). Furthermore, Heckman (2001) argues that it is vital that models account for the obvious differences among individuals, since heterogeneity plays a considerable role in economic behaviour. More recently, heterogeneous agents models (HAM) have been employed to study a range of complex phenomena across a number of disciplines. In finance, HAMs have been used to explain phenomena such as heavy tails in the distribution of returns, bubbles, and other pricing anomalies that would have been impossible to describe with representative agents (see e.g. Brock and Hommes (1997), Lux and Marchesi (1999), Buckley et al. (2012) and Joëts (2015)).

Although a considerable amount of effort has been expended in finding the roots of pricing anomalies in financial markets, some of these anomalies are left without a fully convincing explanation, mainly due to the use of overly simplistic models. One of these mispricing phenomena is the favourite-longshot bias (FLB).

The FLB is an empirical regularity found in state-contingent claims markets, whereby the average return on likely outcomes is greater than the average return on less likely outcomes. That is, to use sports betting terminology, favourites are underbet and longshots overbet; i.e. the chance of high/low probability events are under-/over-estimated. This market anomaly has been extensively studied in sports betting markets, which have the essential characteristic of interest to us, i.e. being able to extract forecast probabilities about future outcomes from the spending behavior of the crowd (Ma et al. (2016)). However, the importance of the FLB goes well beyond sports betting markets. In particular, given the similarities between sports betting markets and traditional state-contingent claims markets it is not surprising that the FLB has been observed in a variety of such markets (Hodges et al. (2003) for a study of the bias in S&P500 and FTSE100 index futures options, and Wolfers and Zitzewitz (2004) for a discussion on the FLB in prediction markets). Thus, a model that adequately explains the FLB could be used to estimate the magnitude of mispricing in prediction markets under given conditions, allowing suitable adjustments to be made; thus significantly

improving the accuracy of the predictions.

In this paper, we make an important first step towards this by showing a way to accurately reconstruct price curves that account for mispricing caused by the FLB. This model can be reverse engineered to derive the true probability of an event given the biased market prices, consequently improving the forecasting power of prediction markets.

The existing literature proposes a variety of theories to explain the FLB. All these indicate that the bias arises given some assumptions.¹ These theories all lead to qualitatively similar results, but they exhibit two common drawbacks. First, they lack empirical support across different markets. That is, the models are usually only tested on a single market. Second, these theories usually employ a representative agent to model bettors (or, at most, a representative agent beside noise traders). Consequently, these models are insufficiently flexible to explain related phenomena such as the reverse FLB (Woodland and Woodland, 1994, 2003) or the FLB in markets beyond the one studied.

Our goal is to build an appropriately complex model capable of providing a comprehensive explanation of the FLB in its various forms. We achieve this by generalizing the most important theories that try to explain this phenomenon into a HAM with five agent types. The potential for a HAM to explain the FLB is highlighted by the observation that betting markets feature a variety of traders who display significantly diverse behaviours (Rhoda et al., 1999). Furthermore, Crawford and Pendakur (2013) find that a representative agent can only explain two thirds of the variation in consumption behaviour, while using four or five classes of agents can completely rationalise all consumer choices in their data.

We model a fixed-odds market using an agent-based model. We focus on developing a model to explain the FLB in a betting market with a market maker, since these markets have been shown to be those most prone to the FLB. In the model, a bookmaker faces five different types of bettors, each associated with a different behaviour suggested in the literature and modelled using prospect theory. That is, we derive the maximum price (i.e. minimum odds) bettors are willing to accept, depending on the behaviour class to which they belong. **This allows us to build a model in which agents with different behaviours, borrowed from literature explaining the FLB with risk preference (e.g. Ali (1977), Golec and Tamarkin (1998)) and misbelief (e.g. e.g. Snowberg and Wolfers (2010), Gandhi and Serrano-Padial (2015)), participate simultaneously in the market as separate entities, hence enabling us to measure their relative contribution to price formation.**

Most similar to our approach are the works by Chiappori et al. (2012) and Gandhi and Serrano-Padial (2015), who have recently added to the debate by introducing heterogeneity in their models, which focus only on either heterogeneous preferences (Chiappori et al., 2012) or heterogeneous beliefs (Gandhi and Serrano-Padial, 2015). In contrast, the HAM we propose in this paper considers heterogeneity on both beliefs and risk preferences simultaneously. Our results, similarly to theirs, suggest that heterogeneous agents are fundamental to capture the full complexity of the FLB and, more generally, of prediction markets. However, we identify three important contributions that distinguish our paper from previous literature. First, we show that our HAM, unlike any of the existing representative agent models, can explain different degrees of the FLB and the reverse FLB. Indeed, we prove that, in the presence of heterogeneous bettors, the FLB and its negative counterpart can occur regardless of whether the market maker adopts a profit maximisation or a risk minimisation pricing strategy.

Second, we analyse two different pricing strategies the market maker can adopt. Knowing the market maker’s pricing strategy is fundamental to being able to reconstruct the price curve, and a

¹See Ottaviani and Sørensen (2008) for a survey on the FLB in sports betting markets.

necessary first step for more accurate forecasts. We show that using our HAM, the best fit to FLB data from three sports with very diverse degrees and types of FLB is achieved if bookmakers act as risk minimisers, whereas the common assumption is that bookmakers are profit maximisers. The only exception is by Fingleton and Waldron (1999), as they analyse the odds from 1696 races in Ireland in 1993 and reject the hypothesis that bookmakers seek to maximise their expected profit, but also fail to reject the hypothesis that bookmakers are risk minimisers, suggesting that this might be their behaviour. Also, our results agree with recent empirical work by Kopriva (2009) and Feess et al. (2016), who found that, on data from Betfair and the New Zealand Racing Board respectively, bet sizes are significantly decreasing in odds (i.e., volumes on longshots are systematically lower). Our finding is important since this knowledge enables appropriate construction of price curves which can enable us to adjust prediction market forecasts to better reflect the true underlying event probabilities.

Third, we determine the distribution of market prices for different market compositions (i.e. agent proportions) using agent-based simulations. Consequently, by observing the FLB for the three sports betting markets we examine (i.e. tennis win market, baseball win market, and under-over 2.5 goals in football), we are able to find their market compositions. This enables us to reconstruct the price curves given the probability of an outcome and adjust for the FLB, hence improving the accuracy of forecasts derived from this prediction market.

Our results suggest that our HAM is better than representative agent models at describing prices in betting markets. More generally, these findings suggest that a HAM is necessary to fully understand the FLB, supporting the idea that representative agent models cannot describe complex systems in a general way with sufficient accuracy.

The paper is organised as follows. In Section 2 we define the agents and introduce the formal model. In Section 3 we derive and discuss the theoretical results, proving that a HAM can explain the FLB under different market conditions. Section 4 focuses on empirical findings, such as the estimation of equilibrium price curves and market compositions, and the comparison with existing theories. In this section we also perform a sensitivity analysis of agent classes' contributions to equilibrium prices. We draw conclusions in Section 5.

2. The Model

To model the betting market, we use an agent-based model, in which one of the agents act as a price-setter (the bookmaker) and the others represent traders (bettors). The bookmaker faces many bettors, and sets the odds in the market depending on the bettors' actions, according to one of two possible pricing strategies. Bettors have heterogeneous beliefs and attitudes towards risk depending on one of the five classes to which they are assumed to belong. As a result, the value they assign to gambles differs depending upon the *behaviour class* to which they belong. We derive the value they assign to gambles from prospect theory by computing the maximum price they are willing to pay to bet on an event.

We consider a time-limited market in which there is only one event with two possible outcomes A and B, occurring with probabilities p_A and $p_B = 1 - p_A$ respectively. Let π_A and π_B be the prices of two contracts that pay one pound if the corresponding outcome occurs, and zero otherwise (Arrow-Debreu securities). Hence, prices are bounded by $0 < \pi_i \leq 1$, where $i \in \{A, B\}$. The boundary value $\pi_i = 0^+$ implies a possible infinite return on the bet amount and $\pi_i = 1$ implies no possible returns or, from the bookmaker's point of view, that bets are not accepted.

2.1. Heterogeneous Bettors and Subjective Fair Prices

We model six different types of agents, namely the bookmaker and five behaviour classes of bettors. Each behaviour class is associated with a set of beliefs and risk attitude such that, given a probability p_i , each agent considers $\pi_i^s(p_i)$ to be the fair price for buying the i -th ticket. That is, they are willing to bet on the i -th outcome if and only if $\pi_i < \pi_i^s(p_i)$, where $\pi_i^s(p_i)$ is referred as the *subjective fair price*. We assume that agents do not possess any information regarding the market, the bookmaker, or other participants. Specifically, they do not know the market composition, other agents' utility functions, or the bookmaker's strategy. Therefore, their decision making process (i.e., whether to bet or not) is solely based on their utility function, which is derived below for each bettor type.

To derive bettors' subjective fair price functions we use prospect theory, because it allows us to derive all the agents' subjective fair price functions from the same equation. Also, prospect theory has shown to be a good estimator of bettors' preferences in sports betting (Jullien and Salanié, 2000).

In more detail, we define functions for both attitude towards risk and probability weighting. Following Tversky and Kahneman (1992), we represent the value function $v(x)$ as:²

$$v(x) = \begin{cases} x^\alpha & \text{if } x \geq 0 \\ -(-x)^\alpha & \text{if } x < 0 \end{cases} \quad (1)$$

Then, we use Prelec's function (Prelec, 1998) as the probability weighting function:

$$w(p) = e^{-[-\ln(p)]^\beta} \quad (2)$$

This means that each agent associates a (subjective) probability $w(p_i)$ to the i -th outcome, with a corresponding potential profit of $1 - \pi_i$, and therefore expects to lose π_i with probability $w(1 - p_i)$. Combining the two, we can write the utility function as follows:

$$u(\pi_i, p_i) = w(p_i)v(1 - \pi_i) + w(1 - p_i)v(-\pi_i) \quad (3)$$

At equilibrium, every agent must be indifferent between trading and not trading, so we find the subjective fair price by setting $u(\pi_i, p_i) = 0$. There is no generic closed-form solution, but specific solutions for each class of agents can be found after setting α and β . Therefore, we define five different agent classes by assigning them specific combinations of values for α and β , deriving the subjective fair price functions from Eq. (3). We do so for all the agents except the random bettors, who do not seek utility maximisation, and are detailed below. Next, we describe the 5 specific trader classes in more detail (derivations of the fair price functions used are given in the Appendix 5).

The first class, random bettors, represents noise traders, and is modelled in a similar manner to that conducted by Shin (1992). Noise traders are the only bettors who do not have subjective fair prices: they attach probability 1 to an outcome randomly picked from the uniform distribution $\text{Unif}\{1, i, \dots, N\}$ (where N is the number of outcomes). In our case, they bet on either A or B with probability $p = 0.5$ each. These agents can be seen as bettors who do not possess good information or bet without taking it into account. For example, one can think of these traders as betting on

²We set $\lambda = 1$ instead of the suggested value of $\lambda = 2.25$, as we want to model informed bettors as unbiased, risk-neutral traders. By setting $\lambda > 1$ instead, informed bettors' subjective price function would be concave, precluding risk-neutrality.

their favourite team in sports betting or investors hedging their position on physical commodities in the options market (Buckley and Long, 2015).

The second class we introduce is the informed bettor class. The literature often defines informed traders as those who have perfect information about the future state of nature. This characterisation, introduced in sports betting by Shin (1991), assumes that insider traders know the outcome of an event with probability 1. However, it could be argued that this is very unlikely to be the case in practice and the assumption has been criticised. Consequently, we assume that informed agents are expert traders, and they are informed in the sense that they have access to all the public information on the event and are capable of processing it without bias, as assumed by Sauer (1998). Under these premises, they can be considered *professionals*. Thus, they are risk-neutral agents who know the true probabilities associated with the event, implying $\alpha = \beta = 1$. As a result, their subjective fair prices correspond exactly to the true probabilities:

$$\pi^s(p) = p \tag{4}$$

The third behaviour class models misperceptions of probability, and has been extensively described in the literature. Most notably, Snowberg and Wolfers (2010) show that, under the representative-agent assumption, misperceptions are more likely to explain the FLB in horse racing than a risk-loving behaviour. We follow them in modelling these agents as risk-neutral bettors ($\alpha = 1$) who weight probabilities with Prelec's function with a coefficient $\beta = 0.928$. Although β might seem too close to 1, this is sufficient to distinguish misperceiving bettors from informed bettors. Then, the obtained subjective fair price function is:

$$\pi^s(p) = -\frac{e^{-(-\ln(p))^{0.928}}}{-e^{-(-\ln(1-p))^{0.928}} - e^{-(-\ln(p))^{0.928}}} \tag{5}$$

The last two behaviour classes represent agents who are not risk neutral.

Risk aversion is found to be one of the behaviours of bettors by Rhoda et al. (1999). Furthermore, risk aversion is a standard assumption in most economic and financial models. Consequently, considering this behaviour class is essential when developing a comprehensive, general model. To determine the parameters for their utility function we follow Gonzalez and Wu (1999). They conduct an experiment in which participants are asked to take gambles; finding that players' median risk attitude value is $\alpha = 0.49$ (with a standard deviation of $\sigma = 0.04$). For sake of tractability, we round this number to 0.5, without making any significant change to the function. The agents who belong to this behaviour class are not affected by misperceptions of probability, hence $\beta = 1$

Thus, choosing $\alpha = 0.5$ to model their risk attitude, their subjective fair price function is:

$$\pi^s(p) = \frac{p^2}{1 - 2p + 2p^2} \tag{6}$$

The last behaviour class we model represents risk-loving agents. Risk-seeking behaviour is frequently employed to model representative agents in sports betting markets because this behaviour alone can cause the FLB. We model this behaviour class with a risk-seeking value of $\alpha = 2$ but assume no misperceptions. Since α is the exponent of the power function, setting $\alpha = 2$ gives us two possible solutions. One of them is negative for $0 \leq p < 0.5$ and discontinuous in $p = 0.5$, with limits of minus and plus infinity as p approaches 0.5 from left and right, respectively. This is unrealistic and consequently we employ the other solution to model the subjective fair price function for risk-loving agents:

$$\pi^s(p) = \frac{p - \sqrt{p - p^2}}{2p - 1} \tag{7}$$

which is defined for $p \in [0, 0.5) \cup (0.5, 1.0]$ and has a removable discontinuity at $p = 0.5$. However, as shown in Section 4, this is not an issue in computing subjective fair prices.

2.2. The Betting Process

We model the betting mechanism as an iterative process, in which the bookmaker can adjust prices depending on its utility function and on bettors' actions. In order to reproduce the time-limited nature of sports betting markets, this process is repeated until equilibrium is reached or an arbitrary number T of rounds are played. The betting mechanism can be described by three stages:

STAGE 1. Nature chooses the true probabilities $\{p_A, p_B\}$ for the outcomes $\{A, B\}$ and chooses the market composition distribution. That is, Nature fixes the proportion of each behaviour class participating in the market.

STAGE 2. The bookmaker sets a pair of prices $\{\pi_A, \pi_B\}$ and allows the agents to bet on either A or B.

STAGE 3. An arbitrarily large number N of bettors is chosen to participate in the market. The number of agents belonging to each class is determined by the market composition distribution selected in Stage 1. All the bettors see the prices and compute their corresponding utility. This process has three steps. First, bettors compute their subjective fair price on all the outcomes according to their subjective fair price function. Second, they compute their utility for betting on each outcome and the utility for not betting. Last, they perform the action that maximises their utility (i.e. betting on A, on B, or not betting).

Note that the market composition (i.e., the distribution of bettors) represents only potential bettors. That is, since agents have the possibility to choose not to bet, the initial market composition chosen at Stage 3 does not necessarily reflect the distribution of classes that actually bet throughout the market. This allows the model to replicate the nature of betting markets, in which prices fluctuate depending on bettors choices. Also, since price is path dependent, the initial market composition (i.e., the potential distribution of bettors) is more indicative than the final one, because it allows to understand what is the necessary starting condition to have given prices at the end.

Finally, modelling the betting process in this way allows our HAM to replicate two essential features of betting markets, without making unnecessary assumptions about bookmaker's and agents' information and behaviour. First, prices are allowed to change at each betting round, and consequently bettors get different odds depending on when they trade. Second, volumes are not evenly distributed over the two outcomes (see Section 4), and bet sizes on longshots are consistently lower, as observed in recent work by Kopriva (2009) and Feess et al. (2016).

2.3. Price Setting

Depending on the bookmaker's selected strategy, the process has different equilibria (i.e. prices converge to different values). In what follows, we first introduce the general setting in which the bookmaker fixes the odds. We then analyse two possible pricing strategies s/he can adopt, namely profit maximisation and risk minimisation. Profit maximisation has often been assumed to be the only strategy available to the bookmaker (e.g. Shin (1991, 1992, 1993), Levitt (2004)). However, even if a small subset of bettors can outperform bookmakers in terms of predicting outcomes, it is possible for the bookmaker to lose money in a given game if s/he sets prices to maximise the expected profit.

Let π_i^f be the fair price on the i -th outcome. We define the *fair* price on the i -th outcome as the price of a bet such that $\pi_i = \pi_i^f = p_i$. We assume that the bookmaker sets the fair prices, and that a total amount of money V (volume) is bet on the event. We also denote by V_A and V_B the

volumes on A and B, respectively, hence $V = V_A + V_B$. The bookmaker's expected profit is given by:

$$\mathbb{E}(P) = V - \frac{p_A}{\pi_A^f} V_A - \frac{p_B}{\pi_B^f} V_B \quad (8)$$

therefore:

$$\mathbb{E}(P) = 0 \quad \forall V_A, V_B \quad (9)$$

However, in any single game the bookmaker's actual profit may vary depending upon whether A or B occurs, as follows:

$$P = \begin{cases} V - \frac{V_A}{\pi_A^f} & \text{if A occurs} \\ V - \frac{V_B}{\pi_B^f} & \text{if B occurs} \end{cases} \quad (10)$$

Without loss of generality, we normalise volumes by setting $V = 1$, so that V_A and V_B become the percentage of money, or relative volume, bet on A and B. Now, it is clear that the bookmaker has a possibility to incur losses every time that $V_A \neq \pi_A^f$, which also implies $V_B \neq \pi_B^f$. That is, if $V_A > p_A$ (consequently $V_B < p_B$) the bookmaker will experience a loss with probability p_A , while in case that $V_A < p_A$ ($V_B > p_B$), s/he will experience a loss with probability p_B .

Next we study the price-setting strategy the bookmaker uses to maximise their utility. We consider two different utility functions, one assuming the bookmaker wants to maximise their expected profit and the other assuming s/he is only concerned to make the same profit regardless of the outcome, as long as the profit is non-negative. Since the bettors are price sensitive, it is not possible to find the prices explicitly as they depend on the volumes, which in turn depend on the choices made by the agents, which in turn are a function of prices. **However we show under both these assumptions, how a model with bettors with heterogeneous behaviours can account for both the FLB and its negative counterpart.**

ASSUMPTION 1. The bookmaker is a profit maximiser (PMB). That is, s/he prefers to maximise their expected profit function rather than minimise profit volatility. Therefore, their utility is $u(p, V_A, V_B, \pi_A, \pi_B) = \mathbb{E}(P) = V - \frac{p_A}{\pi_A} V_A - \frac{p_B}{\pi_B} V_B$.

ASSUMPTION 2. The bookmaker is a risk minimiser (RMB). That is, s/he prefers to minimise profit volatility rather than maximise the expected profit function. Therefore, their utility is $u(V_A, V_B, \pi_A, \pi_B) = -|\frac{V_A}{\pi_A} - \frac{V_B}{\pi_B}|$.

The latter utility function has a global maximum for $\frac{V_A}{\pi_A} = \frac{V_B}{\pi_B}$. That is, risk is minimised if the bookmaker's profit is the same regardless of the outcome. By assuming that the bookmaker is willing to make zero profit, we can substitute $V - \frac{V_A}{\pi_A} = 0$ and $V - \frac{V_B}{\pi_B} = 0$ in Eq. (10). Then, by solving it for π_A, π_B , and recalling that $V = 1$, we obtain:

$$\begin{cases} \hat{\pi}_A = V_A \\ \hat{\pi}_B = V_B \end{cases} \quad (11)$$

where $\hat{\pi}_A$ and $\hat{\pi}_B$ are those prices that allow the bookmaker to have a profit of zero whatever the outcome, and we refer to them as risk-neutral prices.

To keep the model as simple as possible, we also assume that bookmakers do not have any knowledge of bettor utilities. Consequently, bookmakers do not take into account how changing prices may affect the absolute betting volumes, and only optimise their profits per bet. This is a possible limitation to our description of a bookmaker, which however allows us to make the assumption that bookmakers do not possess full information about the market. This latter assumption has two advantages. First, it enables us to keep the model simple, which is essential to accurately quantify the contribution of each agent class to market prices. Second, it does not force us to make

any unrealistic assumption about the degree of information that bookmakers possess about their (existing and new) customers, which is not known and, more generally, may vary among different bookmakers.

3. Theoretical Results

In this section we analyse the implications of Assumptions 1 and 2, outlined in Section 2.3. We show how, under the assumption that the market is composed of heterogeneous agents, a profit maximisation or a risk minimisation strategy can result in the FLB. Moreover, we show how both strategies can also account for the reverse FLB, a market anomaly for which the favourite is overbet and the longshot is underbet. This empirical anomaly has been observed by Busche and Hall (1988) in Hong Kong horse racing and by Woodland and Woodland (1994) in US baseball.

There is no closed-form equation for prices set by a profit-maximising bookmaker under the assumption that agents are price-sensitive. Consequently, we prove that both the FLB and the reverse FLB can be obtained in at least one case by using fixed agent proportions. That is, we make use of two special cases of the profit-maximising bookmaker model which make it analytically tractable, and prove that in these cases either the FLB or the reverse FLB exist.

3.1. The favourite-longshot Bias

In a two-outcome market, the FLB exists if the following holds:

$$\frac{p_A}{p_B} > \frac{\pi_A}{\pi_B} \Leftrightarrow p_A > p_B \quad (12)$$

Therefore, by assuming A and B are the most and least likely outcomes (favourite vs. longshot, or $p_A > p_B$), we can formulate three propositions, whose rigorous proofs can be found in the Appendix Appendix.4. The propositions share the assumptions that bettors are price-sensitive, belong to one of the five classes described in Section 2.1 and behave accordingly. In addition, we assume that the market has the structure described in Section 2.2.

PROPOSITION 1. With a profit-maximising bookmaker (PMB), in a market populated only by risk-loving traders with subjective fair prices defined by Eq. (7), there is always a FLB.

PROPOSITION 2. With a profit-maximising bookmaker (PMB), in a market populated only by risk-averse traders with subjective fair prices defined by Eq. (6), there is always a reverse FLB.

The reasoning behind the propositions is as follows: a bookmaker will make a maximum profit if s/he persuades as many people as possible to bet on the outcome that gives the bookmaker the highest expected profit. In a market in which all the traders share the same beliefs, the bookmaker's best strategy is to set the maximum price the agents would pay for betting on the outcome that guarantees the bookmaker the best expected return. For example, say there is a tennis match in which player A has a true probability of winning of $p_A = 0.8$. For risk-loving bettors, subjective fair prices for this event are $\pi_A^s = 0.\bar{6}$ and $\pi_B^s = 0.\bar{3}$. This is the equilibrium for the agents, meaning that they are indifferent between betting on either A or B, or not betting at all. Since all the traders bet on the same outcome we can split the expected profit in two parts. Hence, the expected profit for the bookmaker is $P(A) = 1 - \frac{p_A}{\pi_A^s} \simeq -0.2$ and $P(B) = 1 - \frac{p_B}{\pi_B^s} \simeq 0.4$ if all the bettors bet on A, or B, respectively. This implies that it is sensible for the bookmaker to set prices $\pi_B = \pi_B^s$ and $\pi_A > \pi_A^s$ to ensure all the agents bet on B.³ A similar (but opposite) reasoning, can be employed to prove proposition 2.

³For this example we assume, without loss of generality, that traders are willing to bet even if $\pi_B^s = \pi_B$ rather than the stricter $\pi_B^s < \pi_B$.

PROPOSITION 3. With a risk-minimising bookmaker (RMB), in a market in which relative volumes are not equal to true probabilities, there is always either the FLB or the reverse FLB.

The proof for proposition 3 is straightforward. Let k be the percentage deviation of V_A from the true probability on the same outcome p_A , so that $V_A = p_A(1 + k)$. For example, if $p_A = 0.6$ and the percentage V_A of money bet on A is $V_A = 0.8$, then V_A is 33% greater than p_A , or $k = 0.33$. Therefore, since $V_A + V_B = 1$, we obtain the following system of equations for the relative volumes:

$$\begin{cases} V_A = p_A(1 + k) \\ V_B = p_B - p_A k \end{cases} \quad (13)$$

By substituting these in Eq. (11) we obtain the new risk-neutral prices. Then, we can check when inequality (12) holds, i.e. when FLB exists:

$$\frac{p_A}{p_B} > \frac{p_A + p_A k}{p_B - p_A k} \implies k < 0 \quad (14)$$

Recalling Eq. (11) it follows that, if $k \neq 0$, we have either the FLB or the reverse FLB. More specifically, if $k < 0$ ($k > 0$), so that more bets are placed on the favourite A (longshot B), the (reverse) FLB occurs. That is, whenever more money than is fair (i.e. the percentage of total money bet on an outcome is greater than the probability of such outcome to happen, that is $V_i > p_i$) is bet on the longshot, there is a FLB, and whenever more money than is fair is bet on the favourite, there is a negative FLB.

4. Empirical Analysis

In this section we find those market compositions that best fit historical FLB data from three diverse betting markets. We start by describing the three data sets and the agent-based simulations we use to find equilibria of the betting process described in Section 2.2 using different combinations of agent proportions (i.e., populations of potential bettors). We then present the results obtained from the agent-based simulations and discuss these by comparing them with data from real betting markets. We find strong evidence that bookmakers are more likely to follow a risk-maximisation (cf. profit-maximisation) strategy. In particular, the former fits historical data better, returns more realistic market compositions, and gives a more accurate representation of the FLB.

To decide which agent composition is the closest to reality for each market, we run a simulation for each possible market composition combination. We then compute the mean squared error (MSE) between the equilibrium prices generated by the simulations and historical prices for all the tested market compositions and choose the market composition that minimises the MSE. Furthermore, we compare our models with two of the most widely cited models, namely those of Shin (1993) and Snowberg and Wolfers (2010), and show that our RMB model displays significantly smaller MSEs. We also show that our RMB model exhibits higher adaptability, since the errors do not vary significantly between markets. We argue that these results suggest that our HAM can be regarded as a generalisation of existing models of the FLB.

4.1. Data

We use three data sets in our analysis, involving historical prices on tennis and baseball win bets and under-over 2.5 goals bets in football. The data on tennis and football are publicly available from specialised websites⁴, while the baseball data are supplied in Woodland and Woodland (2003).

⁴www.tennis-data.co.uk and www.football-data.co.uk for tennis and football, respectively.

Table 1
Best Fitting Market Compositions and Respective MSEs Between the Prices Generated and Historical Data

	Tennis		Under-over 2.5		Baseball	
	RMB	PMB	RMB	PMB	RMB	PMB
Random Agents	10%	0%	45%	20%	5%	0%
Risk-loving Agents	0%	5%	5%	20%	5%	0%
Risk-averse Agents	0%	0%	35%	0%	55%	0%
Misperceiving Agents	60%	40%	15%	5%	5%	40%
Informed Agents	30%	55%	0%	55%	30%	60%
MSE	0.00003	0.00034	0.00020	0.00127	0.00017	0.00022

These data have been employed in several studies (e.g. Forrest and Simmons (2008), Hvattum and Arntzen (2010), Forrest and Mchale (2007), Malueg and Yates (2010)). One of the reasons we choose these data is that they relate to two outcome events, since in tennis and baseball a draw is not possible and in a football match the total number of goals scored is always greater or less than 2.5. Most importantly, the data sets present three diverse levels of the FLB. That is, the tennis and baseball win markets are very efficient compared to the under-over 2.5 (which displays a large FLB).⁵ Baseball odds, instead, have been shown to display a negative FLB. Hence, these data enable us to test our model under three different scenarios (small, large, and negative FLB).

The tennis data incorporate the mean value of the odds provided by five different bookmakers on both the competitors in 27,069 matches in the ATP circuit over ten seasons, from January 2005 to November 2014. The football data incorporate a total of 13,013 matches in the seasons between 2005/2006 and 2013/2014, considering mean odds across 30 bookmakers on the four major European football leagues (Italian Serie A, English Premier League, Spanish Liga, and German Bundesliga). The baseball data incorporate odds data on 44,675 Major League Baseball matches between 1978 and 1999. The baseball data differs to that of the football and tennis in that they use odds lines rather than decimal odds, and odds across bookmakers are already aggregated.

The data sets we employ are sufficiently large to test the goodness-of-fit of our model, and are one to two orders of magnitude larger than most of the earlier studies on the topic (other than a few studies which employ data for horse racing, e.g. Snowberg and Wolfers (2010)).

4.2. The Tennis Market

We define the *best market composition* as that distribution of agents (bettors) which fits the data best (i.e. the market composition whose generated prices minimise the MSE between these prices and those displayed in the real market). Table 1 displays the best market composition for the PMB and the RMB models for all the data sets. In tennis, both the PMB and the RMB models give similar results, namely that most of the betting volume comes from informed and misperceiving bettors.

Although at a first sight it may seem strange that 90% to 95% of the total money wagered on tennis is placed by these classes, this is consistent with Forrest and McHale (2005)'s argument. In fact, this and a subsequent paper (Forrest and Mchale, 2007) agree that most tennis bettors can be considered sophisticated for the following reasons: first, it is easier to assess the relative

⁵Cain et al. (2003) examine the FLB in nine win markets from eight different sports, but only the odds on Cricket (although with only 264 observations) seem to be affected by the bias to a greater extent than that found in our under-over 2.5 data set.

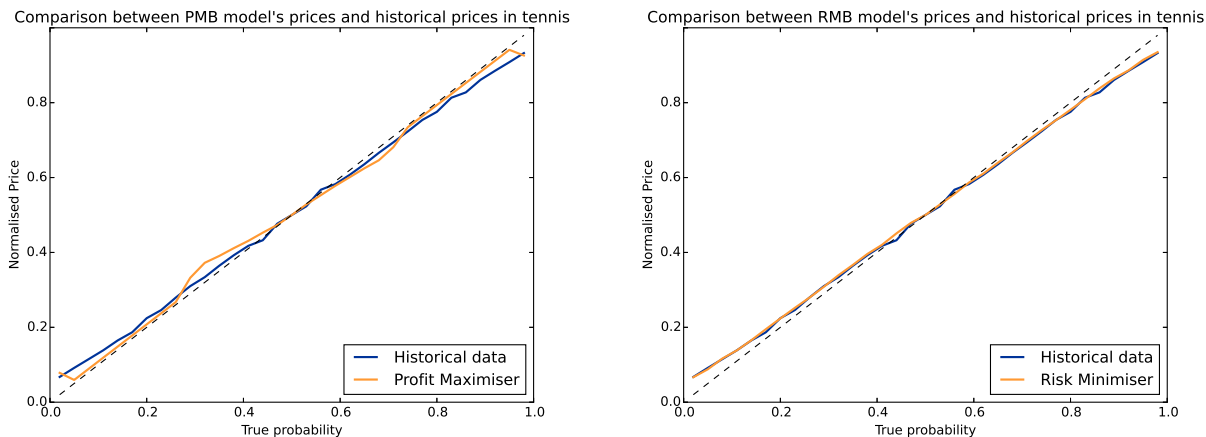


Figure 1: Comparison between equilibrium prices generated by the best market compositions (i.e. those market compositions for which the prices generated are closer to historical prices) in tennis and historical prices. Figure 4.2 represents prices fixed by a profit maximising bookmaker, while figure 4.2 represents prices fixed by a risk minimising bookmaker. The dashed line represents the fair prices.

quality of tennis players, and thus their win probabilities, than it is to assess the relative merits of competitors in other sports (e.g. horse and greyhound racing). This arises because there are only two possible outcomes for a game and because available information such as the ATP rankings already incorporate a strong predictive power (del Corral and Prieto-Rodríguez, 2010). Hence it is easier to become an informed bettor.

Second, in contrast to team sports, there is no significant sentiment betting in tennis. Only a few players can boast a large number of fans, and even these numbers are not comparable to the millions of supporters of major football or basketball teams. Since sentiment bettors place bets regardless of the probability of a team winning, they can be included in the random bettors class. Consequently, in tennis we expect fewer random bettors.

In the tennis betting market, even though both the PMB and the RMB models support similar market compositions, equilibrium prices tell a different story. The equilibrium prices fixed by the RMB give an MSE one order of magnitude smaller than that in the PMB model (i.e. 0.00003 for the RMB and 0.00034 for the PMB). Moreover, by looking at figure 4.2 we note that the prices generated by the PMB are closer to the diagonal, which represents the fair prices, than to the historical prices. By contrast, the equilibrium prices generated by the risk minimising boomaker simulations overlap almost perfectly with the historical data. Therefore, we conclude that the RMB model better explains the prevailing FLB in the tennis win market.

4.3. The Football Market

The under-over 2.5 goals market shows a very different equilibrium price curve than that related to tennis, with a more pronounced FLB. The MSEs are an order of magnitude greater than those obtained using the tennis data.⁶ However, the RMB model still returns a smaller MSE than the PMB. In addition, the prices obtained from the simulation based on a PMB fail to trace the historical price curve properly. In particular, the curve fails to reproduce its peculiar S shape (see figure 2).

⁶Part of the increment of the MSE can be attributed to the higher noise present in the under-over 2.5 data. The smaller number of observations and the shorter range of odds are likely to be the main causes of this noise. In figures 4.3 and 4.3 the effect of the noise in the historical prices curve can be seen clearly.

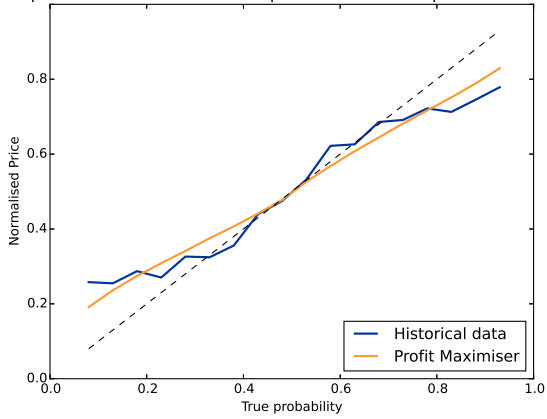
The most interesting observation is that, with this data set, the simulation results suggest very different market compositions when assuming a PMB or a RMB. Previous literature and some of the features of the goal scoring process in football suggest that the RMB model is likely to give the most reasonable market composition. This is likely because, amongst other things, the number of *points* is rather limited (cf. other sports such as tennis or basketball). This means that random events (e.g. players' or referee's errors) can readily make the difference between more or less than 2.5 goals being scored. Furthermore, there is strong evidence of momentum in football (Bittner et al., 2007) implying that, after a team score, they are more likely to score again. Moreover, Armatas et al. (2007) suggest that in-play-only observable variables such as physical condition, coaches' choices, and lapses in concentration are key features in determining the number of goals scored by each team. The most significant evidence for the randomness of the goal scoring process comes from the odds on offer in these markets. In our under-over 2.5 market data the average normalised odds value is 2.03, with a standard deviation $\sigma = 0.3$, suggesting that, on average, bookmakers rate both the under and the over 2.5 goals to be almost equally likely. Therefore, we can conclude that goal scoring in football is a highly unpredictable stochastic process. As a result, it would be reasonable to expect many random bettors and few informed individuals to bet on such events.

As shown in Table 1, the risk minimisation model meets these expectations. Assuming that bookmakers follow the RMB model, the simulation suggests a market composition in which 45% of the money is bet by random bettors and no informed bettors play in the market. The PMB model, on the other hand, suggests a market composition including 55% informed bettors and no random bettors. Similarly, the incidence of risk-loving and risk-averse bettors suggested by the RMB appears more realistic than those suggested by the PMB model. In particular, the raw average odds value (i.e. not normalised) is 1.9 ($\sigma = 0.28$), which is rather low compared to other sports, especially those with more than two outcomes. This would suggest that only a few risk-loving agents would participate, while many risk-averse bettors should be attracted to bet. This is reflected in the RMB model's market composition, in which 35% of the total money is bet by risk-averse agents and only 5% comes from risk lovers. By contrast, the PMB model suggests a market with no risk-averse bettors and where 20% of the volume derives from the bets of risk lovers. Therefore, we conclude that the risk minimisation model is the one that gives the results that are most plausible for the under-over 2.5 market.

4.4. *The Baseball Market*

The RMB model again gives more plausible results when applied to baseball data (see Table 1). Although the MSE is similar for both the models (0.00017 for the RMB and 0.00022 for the PMB), the equilibrium prices derived by the PMB model do not reflect the odds derived from historical data. In fact, the simulation results indicate that the equilibrium prices generated by the PMB model display a positive FLB, which is unsupported by historical prices. Even though its level is rather low, this implies that the PMB model fails to give qualitatively good results. Instead, the price curve obtained by the simulation assuming bookmakers are risk minimisers, supports a negative FLB, as displayed in historical prices. The market composition derived by the simulation suggests that 35% and 55% of the money is bet by informed and risk-averse agents, respectively. The proportion of informed bettors may appear low given that baseball bettors are generally considered sophisticated (Woodland and Woodland, 1994). However, this can be explained by the odds on offer in these markets. Indeed, across the 44,675 games, the odds range between 1.3 and 4.1, and 90% of them lie between 1.59 and 2.7. With such low odds, it is no surprise that risk-averse bettors are attracted to the market.

Comparison between PMB model's prices and historical prices in under-over



Comparison between RMB model's prices and historical prices in under-over

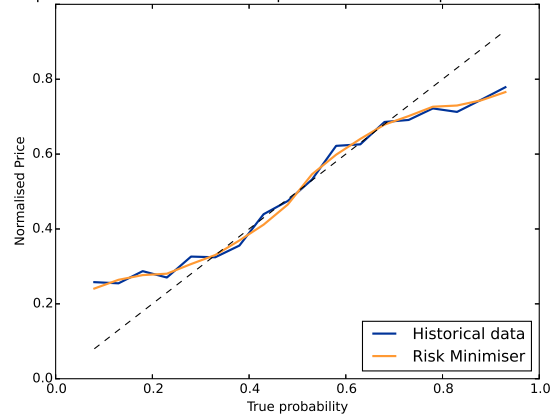
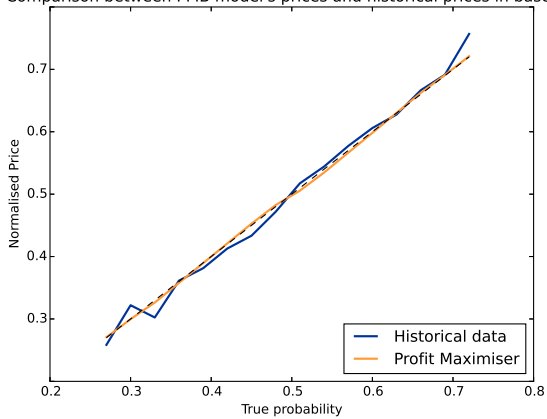


Figure 2: Comparison between equilibrium prices generated by the best market compositions (i.e. those market compositions for which the prices generated are closer to historical prices) in under-over 2.5 and historical prices. Figure 4.3 represents prices fixed by the profit maximising bookmaker, while figure 4.3 represents prices fixed by the risk minimising bookmaker. The dashed line represents the fair prices.

Comparison between PMB model's prices and historical prices in baseball



Comparison between RMB model's prices and historical prices in baseball

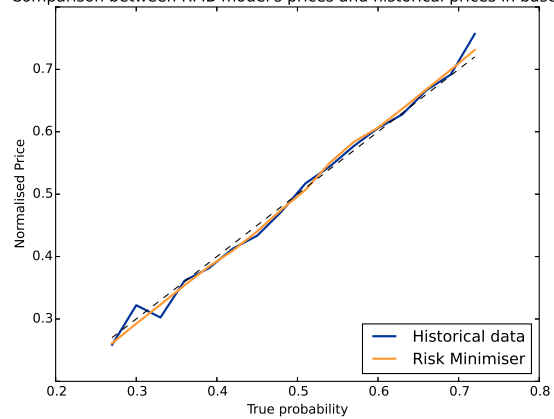


Figure 3: Comparison between equilibrium prices generated by the best market compositions (i.e. those market compositions for which the prices generated are closer to historical prices) in baseball and historical prices. Figure 4.4 represents prices fixed by the profit maximising, while figure 4.4 represents prices fixed by the risk minimising. The dashed line represents the fair prices.

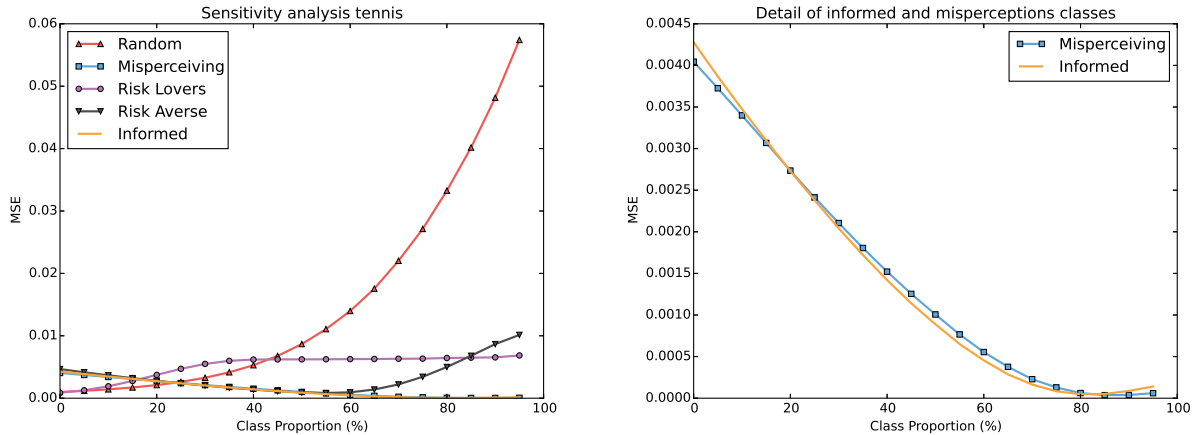


Figure 4: Average MSE in tennis betting markets associated with different proportions of agents from the five behaviour classes. Figure 4.5 shows in detail the curves for misperceiving and informed agents.

4.5. Sensitivity Analysis

In Section 4.2 we presented empirical evidence that suggests that the average bookmaker is more likely to adopt a risk minimising (cf. profit maximising) pricing strategy. In this section we show, through a sensitivity analysis, that every agent behaviour class plays a significant role in shaping equilibrium prices. Consequently, we examine the robustness of employing representative agents, and suggest that a HAM is necessary to explain the FLB.

The above analysis indicates that certain types of bettors are attracted to particular markets. All the classes contribute to shaping price curves (as can be seen in figure 4, 5, and 6), but they simply bet in the markets which are best suited to maximising their utility. This is especially true for risk-loving agents, who are not present in the best fitting market composition in tennis and baseball, and only account for the 5% of the money bet on under-over 2.5. That can be intuitively explained by the fact that odds in those markets are rarely high, mainly because there are only two possible outcomes. Thus, it seems reasonable that risk lovers prefer to bet on other sports. For instance, in horse racing, strong favourites have odds similar to the average odds in the sports we examine. In addition, in horse racing it is possible to find bets paying off up to 1000 times the amount bet. In particular, risk-loving bettors are, by definition, inclined to bet on high odds, since this would return a higher utility (as described in Eq. (7)).

We analyse the specific contribution of each agent behaviour class under the RMB assumption. This is computed as follows: We fix the proportion of a specific agent class in the market and then we compute the average MSE of all the combinations that include this behaviour class proportion. By doing so, we can measure the contribution of each behaviour class. Then, by changing the proportion of that behaviour class we see how the MSE changes. For example, in tennis, the average MSE of all the market compositions with 0% random bettors is 0.00099, and it increases exponentially, reaching an average MSE of 0.0619 when only random bettors participate in the market. As can be seen in figure 4.5, random bettors are those that affect tennis equilibrium prices the most, increasing the MSE exponentially as they grow in number. Also, we note that misperceiving and informed bettors contribute in a similar degree to tennis equilibrium prices, but their sensitivity is significantly different. This confirms that choosing $\beta = 0.928$ as the weighting function's parameter is sufficient to model misperceptions. Figure 4.5 provides another important observation, showing that the misperceptions behaviour class would be the best to describe the FLB in a representative agent model in a fixed-odds market with a RMB. This confirms the findings of

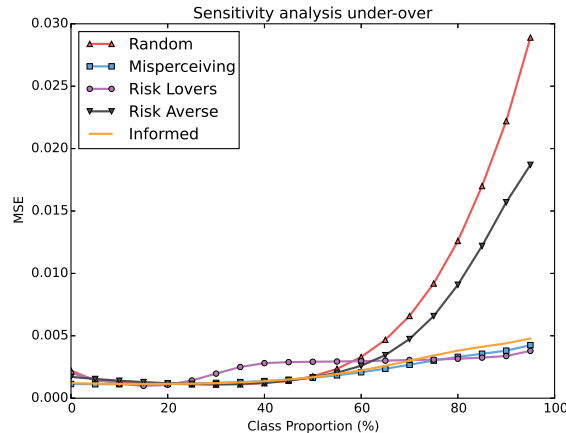


Figure 5: Average MSE in under-over 2.5 betting markets with different proportions of agents from the five behaviour classes.

Snowberg and Wolfers (2010).

Figure 5 demonstrates that the random agents’ MSE curve in the under-over 2.5 market is similar to that for the tennis market, but this time random agents are more likely to play a significant role in price shaping. In fact, when random agents account for 30%-45% of the money bet in the market this produces the lowest MSEs. Other low MSEs can be found when misperceiving or informed bettors are almost or completely absent, suggesting that these categories are not attracted to under-over 2.5 betting markets.

Finally, we find that risk lovers are the class of bettors that make the best representative agent in under-over 2.5 betting markets. It is interesting to note that misperceptions and risk-love, which define the two most widely employed bettor types in the literature, are the most plausible explanations with a representative agent approach in both tennis and under-over 2.5. In this sense, our work supports the claims of Snowberg and Wolfers (2010). They compare misperceptions and risk-love explanations for the FLB in parimutuel horse racing markets and conclude that misperceptions better describe the bias.

However, as a result of our sensitivity analysis, we find that the best type of representative agent changes depending on the market. A further confirmation of the necessity of heterogeneous agents is provided by bettors’ sensitivity in baseball. While risk-averse bettors make up the majority of the market, accounting for 55% of the money bet, the best representative agent would be an informed bettor, with an average MSE roughly thirty times smaller than a risk-averse agent (figure 4.5). However, choosing informed agents to represent all the bettors would make the negative FLB impossible to explain.

To conclude, our sensitivity analysis provides strong evidence that HAMs are necessary to fully rationalise the FLB, thus enabling one model to explain the bias in different markets.

4.6. Comparison with the models by Shin and Snowberg and Wolfers

In Section 4.2 we showed that the RMB model outperforms the PMB in all the three data sets we use, generating smaller MSE and suggesting more plausible market compositions. In this section, we compare it with the models developed by Shin (1993) and Snowberg and Wolfers (2010). Snowberg and Wolfers (2010) build two representative agent models based on previous literature related to risk attitude and misperceptions of probabilities. They conclude that the latter better models the FLB. However, they do not account for information asymmetry as a possible cause of the FLB.

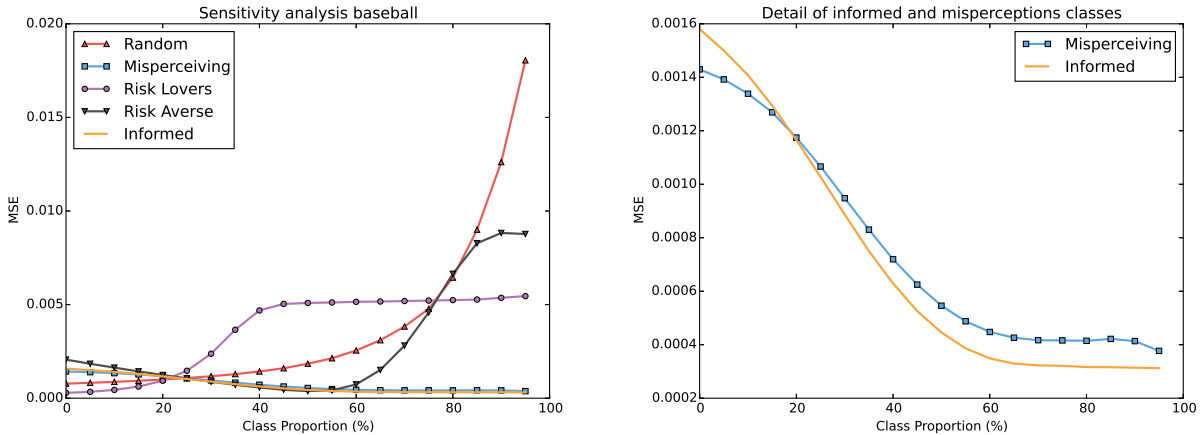


Figure 6: Average MSE in tennis betting markets with different proportions of agents from the five behaviour classes. Figure 4.5 shows in detail the curves for misperceiving and informed agents.

Table 2
Comparison Between Models to explain FLB in Literature and Our RMB Model

	Mean Squared Error		
	Tennis	Under-Over 2.5	Baseball
RMB Model	0.00003	0.00024	0.00017
Snowberg and Wolfers (2010)	0.00020	0.00276	0.00046
Shin (1993)	0.00015	0.00158	0.00028

Consequently, we also compare our results with those derived from the FLB model developed by Shin (1993). In fact, his is one of the most cited works in the information asymmetry stream and a foundation for many other papers studying market efficiency in sports betting (e.g. see Bruce et al. (2012) and Smith et al. (2006)).

We compute the MSE between our data and the prices generated by the models of Shin and Snowberg and Wolfers and compare these MSEs with those obtained by using the RMB model (see table 2). We note that all three models give a similar MSE on baseball, but the RMB model performs significantly better on the data from tennis and under-over 2.5. This suggests that our HAM is highly adaptable and can describe different markets without significantly decreasing its accuracy. By contrast, the representative agent models are tailored to specific markets, so they often only work well in conditions analogous to those assumed in the original model. This, of course, should not be surprising. Indeed, in the last decades, many economists have criticised the representative agent approach more generally because of its lack of flexibility. Another common criticism is that a representative agent approach inevitably misses some key features of complex systems, since it completely neglects interactions among different types of agents.

The MSEs of the models displayed in Table 2 show that the RMB model explains historical prices far better than the other two models. This clearly shows the advantages of adopting a heterogenous agents model to describe the FLB. In fact, we note that by using Snowberg and Wolfers (2010) we implicitly assume that all bettors misestimate probabilities in the same way, and that they are perfectly neutral towards risk. However, this is not the behaviour that it is found in many studies exploring how people gamble. In fact, after Kahneman and Tversky (1979) introduced prospect theory, many papers focused on measuring people’s preferences for risk, seldom finding them to be neutral.

Similarly, applying Shin (1993) to different markets results in predictions of high incidence of insider trading. This is very surprising given that Shin’s insiders are assumed to know exactly the outcome of an event. Analysis of our data using the Shin model suggests that 5% and 10% of the money bet on baseball and tennis respectively are from bettors who know the exact outcome of the game. This percentage rises dramatically to 46% in the under-over 2.5 market. These values seem unrealistically high, especially considering the highly stochastic nature of the goal scoring process, as explained in Section 4.2. Interestingly, we find that the percentage of money bet by Shin’s insiders is inversely proportional to the percentage of money bet by informed bettors in the RMB model. This reflects the different views of the market represented in these models. Shin’s insiders make the market more inefficient, while in our model informed bettors contribute to increase market efficiency.

The comparison between the results derived from our HAM and those derived from the Shin and Snowberg and Wolfers models provide a clear picture of the important difference between these representative and heterogeneous agent models. Snowberg and Wolfers (2010) build their model to explain the FLB in the horse race pari-mutuel market, which has a different structure than that in fixed-odds markets. Thus, the parameters chosen will inevitably no longer be optimal for the markets we consider. In addition, Shin makes the assumption that bookmakers set odds such that they do not make any profit or loss, thus making it difficult to compare with our model.

To overcome this problem, we run new simulations by calibrating both Shin’s and Snowberg and Wolfers’ models for each data set⁷. We find that errors remain of the same magnitude⁸. However, we notice interesting changes in the models’ key parameters. Shin’s measure of the incidence of insiders does not change significantly and remains high especially in the under-over market; his calibrated model suggesting that 45% of the money is bet by insider bettors. His calibrated model also suggests that 5% and 10% of betting in baseball and tennis, respectively, is undertaken by insiders. However, we find important changes in Snowberg and Wolfers’ model. Prelec function’s β values in this model are considerably lower (0.87 and 0.64 for tennis and under-over 2.5, respectively) than the value they proposed in their paper, that is $\beta = 0.928$. This implies greater misperceptions of probability in these markets compared to horse racing. This result is in line with what we expected for the under-over 2.5 market. However, it is at odds with our expectations for tennis, where the FLB is lower. In addition, this is out of line with the behaviour of tennis bettors, as discussed in Section 4.2.

For baseball, we find an optimal value of $\beta = 1.06$ using Snowberg and Wolfers’ model. A value of $\beta > 1$ is necessary to lead to the reverse FLB in a representative agent model where the representative bettor is one who misperceives probabilities (since small probabilities are underweighted and large probabilities are overweighted). However, a value of $\beta > 1$ is out of line with theoretical work on misperceptions of probabilities, including the work of Prelec himself, which suggests that $0 < \beta < 1$ must hold (Prelec, 1998). This provides further evidence that a representative agent model cannot explain all the aspects of the FLB.

⁷For the model by Snowberg and Wolfers we optimise the free parameter β in Prelec’s probability weighting function. For Shin’s model, we allow the bookmakers to seek positive profit, setting the margin they are willing to achieve to the average margin found in the historical data sets

⁸The greatest relative change in MSE values is for Snowberg and Wolfers’ model in tennis (decreasing from 0.00020 to 0.00012), while the greatest absolute change occurs in Snowberg and Wolfers’ model for under-over 2.5 (decreasing from 0.00276 to 0.00240)

5. Concluding Remarks and Future Work

In this paper, we seek to understand the factors that lead to a persistent pricing anomaly (FLB), with the overall aim of demonstrating the value of HAMs for understanding the true causes of such phenomena. We argue that such an approach offers the prospect of significantly improving the accuracy of forecasts derived from prediction markets. Specifically, we generalise existing theories designed to establish the causes of the FLB, and use these to develop a HAM. Our results show that a HAM is necessary to fully understand this important pricing anomaly, supporting the idea that representative agent models cannot describe complex systems in a general way with sufficient accuracy. Indeed, our comparisons with established models show that those using a representative agent cannot adapt to different markets. Furthermore, we demonstrate that our bookmaker risk minimisation model outperforms two of the most influential FLB models (i.e. Shin (1993) and Snowberg and Wolfers (2010)) in terms of reproducing real prices in three different sports betting markets and, most importantly, in explaining the regular and reverse form of the FLB.

Our results help to confirm the important role which analysing behaviour in sports betting markets can perform for understanding behaviour in other state-contingent claims markets. The importance of understanding the behaviours that cause pricing anomalies lies in that knowing the causes enables us to account for such pricing anomalies beforehand. This is fundamental in a number of contexts way beyond sports betting, such as prediction and financial markets.

Our results also suggest that our HAM is better than models based on representative agents at describing betting markets, and, most importantly, our findings provide new evidence that HAMs provide a robust way of effectively modelling pricing anomalies that are caused by heterogeneous behaviours of agents. Our results, therefore, offer the prospect of being able to appropriately adjust the prices that emerge in prediction markets to correct for pricing anomalies, thus leading to more accurate forecasts.

We identify two main lines for future work that will enable us to have a more powerful forecasting tool. First, we believe it is particularly important to extend this model beyond two-outcome events, and fixed-odds markets. This would be especially useful for many internal prediction markets, which consider multi-outcome events, and political elections. Second, in future work, we would like to extend our HAM to better forecast the true probabilities of events given the corresponding market prices. For instance, knowing the circumstances of a particular prediction market (e.g., the range of odds associated with the various potential outcomes) will enable us to estimate the proportions of different agent classes likely to be attracted to the market, and consequently allow us to estimate (and thus adjust for) the degree of FLB displayed in the market prices.

Appendix A : Subjective Fair Prices Derivation

In this section we derive the subjective fair prices function shown in Section 2.1. Recalling the utility function for the agents

$$u(\pi, p) = w(p)v(1 - \pi) + w(1 - p)v(-\pi) \quad (.1)$$

where the value function is

$$v(x) = \begin{cases} x^\alpha & \text{if } x \geq 0 \\ -(-x)^\alpha & \text{if } x < 0 \end{cases} \quad (.2)$$

and the probability weighting function is

$$w(p) = e^{-[-\ln(p)]^\beta} \quad (.3)$$

we derive each subjective fair price function by varying the values of α and β .

Appendix.1. Informed Bettors

The informed bettors are defined by having no misperceptions of probabilities and being risk-neutral, hence $\alpha = 1$ and $\beta = 1$. By substituting such values in equation (.1), we obtain:

$$u(\pi, p) = (1 - \pi)e^{-[-\ln(p)]} - (\pi)e^{-[-\ln(1-p)]} = (1 - \pi)p - \pi(1 - p) \quad (.4)$$

Since at equilibrium $u(\pi, p) = 0$ must hold, then $(1 - \pi)p - \pi(1 - p) = 0$ which, solved by π , gives $\pi = p$.

Appendix.2. Misperceiving Bettors

Misperceiving bettors are characterised by being risk-neutral and having a coefficient $\beta = 0.928$. Hence their utility function is:

$$u(\pi, p) = (1 - \pi)e^{-[-\ln(p)]^{0.928}} - (\pi)e^{-[-\ln(1-p)]^{0.928}} = 0 \quad (.5)$$

By expanding and collecting in terms of π the equation becomes:

$$e^{-[-\ln(p)]^{0.928}} + \pi(-e^{-[-\ln(p)]^{0.928}} - e^{-[-\ln(1-p)]^{0.928}}) = 0 \quad (.6)$$

By subtracting $e^{-[-\ln(p)]^{0.928}}$ from both sides and dividing both sides by $-e^{-[-\ln(p)]^{0.928}} - e^{-[-\ln(1-p)]^{0.928}}$ the subjective fair price function can be rewritten as:

$$\pi = \frac{e^{-[-\ln(p)]^{0.928}}}{-e^{-[-\ln(p)]^{0.928}} - e^{-[-\ln(1-p)]^{0.928}}} \quad (.7)$$

Appendix.3. Risk-loving Bettors

Risk-loving agents have no misperceptions of probabilities, and a risk coefficient of $\alpha = 2$. Hence their utility function is:

$$u(\pi, p) = (1 - \pi)^2 e^{-[-\ln(p)]} - (\pi)^2 e^{-[-\ln(1-p)]} = (1 - \pi)^2 p - \pi^2(1 - p) = 0 \quad (.8)$$

By expanding and collecting in terms of π we can rewrite the equation as $p - 2\pi p + \pi^2(2p - 1) = 0$. Then, it is possible to divide both sides by $2p - 1$ and subtract $\frac{p}{2p-1}$ from both sides, obtaining $\pi^2 - \frac{2\pi p}{2p-1} = -\frac{p}{2p-1}$. By adding $\frac{p^2}{2p-1}$ to both sides it is possible to rewrite the left-hand side of the

equation as a square, and consequently to take the square root of both sides, which gives the two following solutions:

$$\pi - \frac{p}{2p-1} = \pm \sqrt{\frac{p^2}{(2p-1)^2} - \frac{p}{2p-1}} \quad (.9)$$

This equation can be reduced to:

$$\pi = \frac{p \pm \sqrt{p-p^2}}{2p-1} \quad (.10)$$

The solution $\pi = \frac{p+\sqrt{p-p^2}}{2p-1}$ goes to minus and plus infinity for p approaching $\frac{1}{2}$ from left and right, respectively. Therefore, this is not a realistic solution and cannot be considered as the subjective fair price function. Hence, the subjective fair price function for risk-loving agents is $\pi = \frac{p-\sqrt{p-p^2}}{2p-1}$, which is defined for $p \in [0, 0.5) \cup (0.5, 1.0]$ and has a removable discontinuity in $p = \frac{1}{2}$.

Appendix.4. Risk-averse Bettors

Risk-averse bettors are characterised by the values $\alpha = 0.5$ and $\beta = 1$. Consequently, their utility function is

$$u(\pi, p) = (1 - \pi)^{\frac{1}{2}} e^{-[\ln(p)]} - (\pi)^{\frac{1}{2}} e^{-[\ln(1-p)]} = (1 - \pi)^2 p - \pi^2 (1 - p) = 0 \quad (.11)$$

By expanding and rearranging the terms it is possible to rewrite the equation as $2\sqrt{-\pi(\pi-1)p(p-1)} = -\pi(p-1)^2 - p^2(1-\pi)$. Then, it is possible to square both sides to obtain $-4\pi(\pi-1)p^2(p-1)^2 = (-\pi(p-1)^2 - p^2(1-\pi))^2$. By expanding and collecting again in terms of π , we obtain the following:

$$\pi^2 + \frac{\pi(4p^4 - 4p^3 + 2p^2)}{-4p^4 + 8p^3 - 8p^2 + 4p - 1} = \frac{p^2}{-4p^4 + 8p^3 - 8p^2 + 4p - 1} \quad (.12)$$

Then, we add

$$\frac{(4p^4 - 4p^3 + 2p^2)^2}{4(-4p^4 + 8p^3 - 8p^2 + 4p - 1)^2} \quad (.13)$$

to both sides, and rewrite the left-hand side as a square, as follows:

$$\left(\pi + \frac{(4p^4 - 4p^3 + 2p^2)^2}{2(-4p^4 + 8p^3 - 8p^2 + 4p - 1)^2}\right)^2 = \frac{p^2}{-4p^4 + 8p^3 - 8p^2 + 4p - 1} + \frac{(4p^4 - 4p^3 + 2p^2)^2}{4(-4p^4 + 8p^3 - 8p^2 + 4p - 1)^2} \quad (.14)$$

The right-hand side is 0. Therefore, by taking the square root of the left-hand side, the equation becomes:

$$\pi = -\frac{(4p^4 - 4p^3 + 2p^2)^2}{2(-4p^4 + 8p^3 - 8p^2 + 4p - 1)^2} \quad (.15)$$

which can be written as follows:

$$\pi = \frac{p^2}{2p^2 - 2p + 1} \quad (.16)$$

Appendix B: Proof of Propositions 1-3

Appendix.5. Proof of Proposition 1

Let us assume that a bookmaker sets prices π_A and π_B for the two possible outcomes of a competitive event, which have probabilities p_A and p_B , respectively. Let us assume that V is the total amount of money bet on the event, and that V_A and V_B represent the amount of money bet on the outcomes A and B, respectively.

PROPOSITION 1. With a profit-maximising bookmaker (PMB), in a market populated only by risk-loving traders with subjective fair prices defined by Eq. (7), there is always a FLB.

PROOF To maximise their profit, the bookmaker needs all the agents to bet on the outcome that will assure the bookmaker the maximum profit. Since the market is populated only by bettors who share the same utility function and beliefs, all the bets will be wagered on a single outcome, either A or B. Therefore, either $V_A = V$ or $V_B = V$ is true. Recalling the derivation in Appendix 5, the risk-loving bettors' subjective fair price is:

$$\pi = \frac{p - \sqrt{p - p^2}}{2p - 1} \quad (.17)$$

By substituting this equation in Eq. (8), we can rewrite the bookmaker's expected profit as follows:

$$\mathbb{E}(P) = V - \frac{p_A(2p_A - 1)}{p_A - \sqrt{p_A - p_A^2}}V_A - \frac{p_B(2p_B - 1)}{p_B - \sqrt{p_B - p_B^2}}V_B \quad (.18)$$

Since there are only two possible outcomes, $p_B = 1 - p_A$, Eq. (.18) can be rewritten as follows:

$$\mathbb{E}(P) = V - \frac{p_A(2p_A - 1)}{p_A - \sqrt{p_A - p_A^2}}V_A - \frac{(1 - p_A)(2(1 - p_A) - 1)}{(1 - p_A) - \sqrt{(1 - p_A) - (1 - p_A)^2}}V_B \quad (.19)$$

Then, at equilibrium, the bookmaker should set risk-lovers' subjective fair price for A if and only if:

$$\frac{p_A(2p_A - 1)}{p_A - \sqrt{p_A - p_A^2}} < \frac{(1 - p_A)(2(1 - p_A) - 1)}{(1 - p_A) - \sqrt{(1 - p_A) - (1 - p_A)^2}} \quad (.20)$$

This can be reduced to $2p_A < 1$, so that the solution is $p_A < \frac{1}{2}$. However, since p_A is the probability associated with the most likely outcome (i.e. $p_A \in (0.5, 1]$), there are no valid solutions to the inequality. Therefore, the bookmaker would lose more money should all the bets be wagered on A instead of B. This suggests that a profit-maximising bookmaker should set the maximum price risk-loving bettors are willing to pay (i.e. the subjective fair price, or $\pi_B = \pi^{(s)}_B$) on B, and a price $\pi_A > \pi^{(s)}_A$ on A. For sake of simplicity, we can assume that the bookmaker sets $\pi_A = \pi^{(s)}_A + \epsilon$, where ϵ has an arbitrarily small positive value. We can then substitute the equations for these prices in inequality (12), which must hold in order to produce the FLB. Then, it is possible to rewrite inequality (12) as:

$$\frac{p_A}{1 - p_A} > \frac{(2(1 - p_A) - 1)(p_A - \sqrt{p_A - p_A^2})}{(2p_A - 1)(1 - p_A - \sqrt{(1 - p_A) - (1 - p_A)^2})} \quad (.21)$$

By subtracting $\frac{p_A}{1 - p_A}$ from both sides and simplifying the expression, the inequality can be rewritten as:

$$\frac{2p_A - \sqrt{p_A - p_A^2}}{p_A - 1} < 0 \quad (.22)$$

The denominator is positive if $p_A > 1$ and negative otherwise, while the numerator is positive only for $p_A > \frac{1}{5}$. In the domain given by $p_A \in [0.5, 1)$, the denominator is always negative and the numerator always positive. Hence inequality (.22) always holds. Therefore, if the bookmaker is a PMB and all the bettors are risk-lovers, the FLB always exists.

Appendix.6. Proof of Proposition 2

PROPOSITION 2. With a profit-maximising bookmaker (PMB), in a market populated only by risk-averse traders with subjective fair prices defined by Eq. (6), there is always a reverse FLB.

To prove proposition 2, we will follow the same reasoning we followed for proposition 1. Let assume all bettors are risk-averse. That is, their subjective fair price function is:

$$\pi(p) = \frac{p^2}{2p^2 - 2p + 1} \quad (.23)$$

Recalling Appendix Appendix.5, the bookmaker should set risk-averse bettors' subjective fair price for A if and only if the following holds:

$$\frac{p_A(2p_A^2 - 2p_A + 1)}{p_A^2} < \frac{(1 - p_A)(2(1 - p_A)^2 - 2(1 - p_A) + 1)}{(1 - p_A)^2} \quad (.24)$$

By expanding and rearranging the terms the inequality can be rewritten as:

$$\frac{(p_A - \frac{1}{2})(p_A^3 - \frac{5}{2}p_A^2 + \frac{3}{2}p_A - \frac{1}{2})}{(p_A - 1)^2 p_A} < 0 \quad (.25)$$

By analysing the sign of the denominator, it is possible to note that, since $p_A \in [0.5, 1)$, the denominator is always positive. Similarly, the term $p_A - \frac{1}{2}$ is always positive. To find the solutions to the inequality, we only need to find the solutions to $(p_A^3 - \frac{5}{2}p_A^2 + \frac{3}{2}p_A - \frac{1}{2}) < 0$. It is possible to rewrite this as $p_A(p_A - 1)(2p_A - 3) < 1$. Clearly, for $p_A \in [0.5, 1)$, the function $f(p_A) = p_A(p_A - 1)(2p_A - 3)$ has a maximum in $f(1) = 0.5$. Hence, it is always less than one, and the inequality always holds. Therefore, when all bettors are risk-averse, for a PMB, it is optimal that the agents bet on A. Following again the procedure employed to prove proposition 1, we rewrite the condition for the FLB to exist in terms of risk-averse bettors' subjective fair prices. However, this time we want to prove that the reverse FLB exists, so we want that $\frac{p_A}{p_B} < \frac{\pi_A}{\pi_B}$:

$$\frac{p_A}{1 - p_A} < \frac{p_A(2(1 - p_A)^2 - 2(1 - p_A) + 2)}{(1 - p_A)(2p_A^2 - 2p_A + 1)} \quad (.26)$$

By expanding, collecting, and rearranging the terms, it is possible to simplify the expression to $(p_A - 1)^2 p_A (2p_A - 1) > 0$. Recalling that $p_A \in (0.5, 1)$, it is clear that all the terms are always positive in this domain. Therefore, we can conclude that in presence of only risk-averse bettors, if the bookmaker is a PMB, there is always a negative FLB.

Appendix.7. Proof of Proposition 3

PROPOSITION 3. With a risk-minimising bookmaker (RMB), in a market in which relative volumes are not equal to true probabilities, there is always either the FLB or the reverse FLB.

PROOF Let k be the percentage deviation of V_A from the true probability p_A on the outcome A, such that $V_A = p_A(1 + k)$. Therefore, since $V_A + V_B = 1$, we obtain the following system of equations for the relative volumes V_A and V_B :

$$\begin{cases} V_A = p_A(1 + k) \\ V_B = p_B - p_A k \end{cases} \quad (.27)$$

Recalling that a RMB sets the prices $\hat{\pi}_A = V_A$ and $\hat{\pi}_B = V_B$, we can substitute these equations of the volumes in $\frac{p_A}{p_B} > \frac{\pi_A}{\pi_B}$ (the necessary condition for the FLB to happen, assuming $p_A > p_B$). Then, we can rewrite it as:

$$\frac{p_A}{p_B} > \frac{p_A + p_A k}{p_B - p_A k} \quad (.28)$$

To solve the inequality for k , we first rewrite the inequality in terms of p_A , as follows:

$$\frac{p_A}{1 - p_A} > \frac{p_A(1 + k)}{1 - p_A - p_A k} \quad (.29)$$

Since $p_A(1 + k)$ represents the relative volume wagered on outcome A, the bounds are $0 \leq p_A(1 + k) \leq 1$, which implies $-1 \leq k \leq \frac{1}{p_A} - 1$.

By rearranging the terms, inequality (.29) becomes $\frac{kp_A}{(p_A-1)(kp_A+p_A-1)} < 0$. Considering the boundaries for p_A and k , we can write a system of inequalities as follows:

$$\begin{cases} \frac{kp_A}{(p_A-1)(kp_A+p_A-1)} < 0 \\ -1 \leq k \leq \frac{1}{p_A} - 1 \\ \frac{1}{2} < p_A \leq 1 \end{cases} \quad (.30)$$

Inequality (.29) has many solutions, but the system has only one, that is $-1 < k < 0$. Therefore, with a RMB, the FLB exists only if the relative volume of money bet on A is greater than the probability of A happening, i.e. $V_A > p_A$. However, following the same reasoning, one can prove that this implies that, if $k > 0$, a negative FLB exists. Furthermore, this implies that a market structured in this way can only be efficient if $k = 0$, i.e. if $V_A = p_A$ and $V_B = p_B$.

Appendix C: Results on Prediction Markets

In this section, we perform an additional analysis on political prediction markets. Our data comprises 3363 observations from the exchange platform PredictIt (www.predictit.org), which are used to reconstruct a price curve for prediction markets. Although our model could be extended to reproduce exchange markets (i.e., by removing bookmakers and allowing agents to short sell), this is beyond the scope of this paper. Therefore, we present our analysis solely with the intention of showing that the results obtained using sports betting data are robust and not the result of overfitting. Our results, displayed in Figure .7, suggest that the RMB can capture the positive FLB exhibited by prediction markets. However, the large value of the MSE (0.0051) indicates that our model is not capable to fully explain the data. Indeed, the MSE computed between historical prices and those results of the simulations, is an order of magnitude greater than those found on the Under-over 2.5 and Baseball markets, and 170 times larger than the error on Tennis. Since our model is not

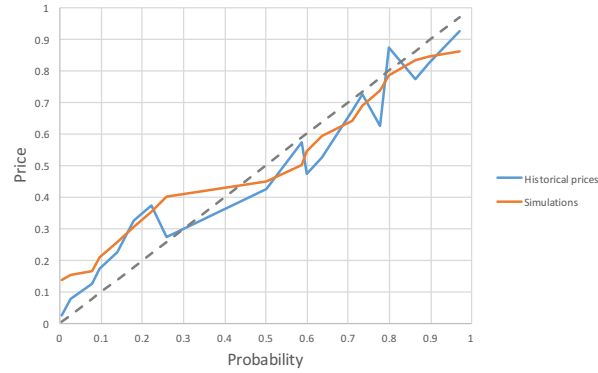


Figure .7: Comparison between equilibrium prices generated by the best market compositions (i.e. those market compositions for which the prices generated are closer to historical prices) in political prediction markets and historical prices. The dashed line represents the fair prices.

suitable to analyse exchange markets, such results were expected, and provide additional evidence that the price curves reconstructed by the RMB for sports betting markets are not a consequence of overfitting, rather, they are a genuine display of the our model’s accuracy. Similarly, the market composition identified by the RMB suggests that, for prediction markets, participants are mostly risk-averse (75%), while only some of them are represented by informed bettors (20%), and the rest are accounted as noise traders (5%). This market composition, however, is difficult to be validated, due to the lack of information about the trader population of prediction markets.

References

- Ali, M. M., 1977. Probability and Utility Estimates for Racetrack Bettors. *The Journal of Political Economy* 85 (4), 803.
- Armatas, V., Yiannakos, A., Sileoglou, P., 2007. Relationship between time and goal scoring in soccer games: Analysis of three World Cups. *International Journal of Performance and Analysis in Sport* 7 (2), 48–58.
- Arnesen, S., Bergfjord, O., 2014. Prediction markets vs polls—an examination of accuracy for the 2008 and 2012 elections. *Journal of Prediction Markets* 8 (3), 24–33.
- Berg, J., Nelson, F., Rietz, T., 2008. Prediction market accuracy in the long run. *International Journal of Forecasting* 24 (285-300).
- Bittner, E., Nussbaumer, A., Janke, W., Weigel, M., 2007. Self-affirmation model for football goal distributions. *Europhysics Letters* 78 (5), 1–5.
- Brock, W., Hommes, C., 1997. A Rational Route to Randomness. *Econometrica: Journal of the Econometric Society* 65 (5), 1059–1095.
- Bruce, A., Johnson, J. E., Peirson, J., 2012. Recreational versus professional bettors: performance differences and efficiency implications. *Economics Letters* 114 (2), 172–174.
- Buckley, W. S., Brown, G. O., Marshall, M., 2012. A mispricing model of stocks under asymmetric information. *European Journal of Operational Research* 221 (3), 584–592.
- Buckley, W. S., Long, H., 2015. A discontinuous mispricing model under asymmetric information. *European Journal of Operational Research* 243 (3), 944–955.
- Busche, K., Hall, C. D., 1988. An exception to the risk preference anomaly. *The Journal of Business* 61 (3), 337–346.
- Cain, M., Law, D., Peel, D., 2003. The Favourite-Longshot Bias, Bookmaker Margins and Insider Trading in a Variety of Betting Markets. *Bulletin of Economic Research* 55:3 (1998), 263–274.
- Chiappori, P., Gandhi, A., Salanié, B., Salanié, F., 2012. From Aggregate Betting Data to Individual Risk Preferences. *Columbia University Working Paper*, n. 13-453.
- Crawford, I., Pendakur, K., 2013. How many types are there? *The Economic Journal* 123, 77–95.
- del Corral, J., Prieto-Rodríguez, J., 2010. Are differences in ranks good predictors for Grand Slam tennis matches? *International Journal of Forecasting* 26 (3), 551–563.
- Fama, E., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *The journal of Finance* 25 (2), 28–30.
- Feess, E., Muller, H., Schumacher, C., 2016. Estimating risk preferences of bettors with different bet sizes. *European Journal of Operational Research* 249 (3), 1102–1112.
- Fingleton, J., Waldron, P., 1999. Optimal Determination of Bookmakers' Betting Odds: Theory and Tests. *Trinity Economic Paper Series* 9 (96).

- Forrest, D., McHale, I., 2005. Longshot bias: insights from the betting market on men's professional tennis. In: Vaughan Williams, L. (Ed.), *Information Efficiency in Financial and Betting Markets*. Cambridge University Press, Ch. 8, pp. 215–230.
- 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.
- Gandhi, A., Serrano-Padial, R., 2015. Does Belief Heterogeneity Explain Asset Prices: The Case of the Longshot Bias. *The Review of Economic Studies* 82 (1), 156–186.
- Golec, J., Tamarkin, M., 1998. Bettors love skewness, not risk, at the horse track. *Journal of Political Economy* 106 (1), 205–255.
- Gonzalez, R., Wu, G., 1999. On the shape of the probability weighting function. *Cognitive psychology* 166, 129–166.
- Heckman, J. J., 2001. Micro Data , Heterogeneity , and the Evaluation of Public Policy: Nobel Lecture. *Journal of Political Economy* 109 (4), 673–748.
- Hodges, S. D., Tompkins, R., Ziemba, W. T., 2003. The Favorite Longshot Bias in S & P 500 and FTSE 100 Index Futures Options: The Return to Bets and the Cost of Insurance. EFA 2003 Annual Conference Paper No. 135; Sauder School of Business Working Paper.
- Hvattum, L. M., Arntzen, H., 2010. Using ELO ratings for match result prediction in association football. *International Journal of Forecasting* 26 (3), 460–470.
- Joëts, M., 2015. Heterogeneous beliefs, regret, and uncertainty: The role of speculation in energy price dynamics. *European Journal of Operational Research* 247 (1), 204–215.
- Jullien, B., Salanié, B., 2000. Estimating Preferences under Risk : The Case of Racetrack Bettors. *Journal of Political Economy* 108 (3), 503–530.
- Kahneman, D., Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47 (2), 263–292.
- Kirman, A. P., 1992. Whom or What Does the Representative Individual Represent? *Journal of Economic Perspectives* 6 (2), 117–136.
- Kopriva, F., 2009. Constant Bet Size? Don't Bet on It! Testing Expected Utility Theory on Betfair Data. CERGE-EI Working Paper series.
- Levitt, S. D., 2004. Why are gambling markets organised so differently from financial markets? *The Economic Journal* 114, 223–246.
- Lux, T., Marchesi, M., 1999. Scaling and criticality in a stochastic multi-agent model of a Financial market. *Nature* 397 (February).
- Ma, T., Tang, L., McGroarty, F., Sung, M.-C., Johnson, J. E. V., 2016. Time is money: Costing the impact of duration misperception in market prices. *European Journal of Operational Research* 255 (2), 397–410.

- Malueg, D. A., Yates, A. J., 2010. Testing Contest Theory: Evidence from Best-of-Three Tennis Matches. *Review of Economics and Statistics* 92, 689–692.
- Ottaviani, M., Sørensen, P. N., 2008. The Favorite-Longshot Bias : An Overview of the Main Explanations. In: Hausch, D. B., Ziemba, W. T. (Eds.), *Handbook of Sports and Lottery markets*. North-Holland, Ch. 6, pp. 83–101.
- Prelec, D., 1998. The Probability Weighting Function. *Econometrica* 66 (3), 497–527.
- Rhoda, K. L., Olson, G. T., Rappaport, J. M., 1999. Risk preferences and information flows in racetrack betting markets. *The Journal of Financial Research* XXII (3), 265–285.
- Sauer, R. D., 1998. The Economics of Wagering Markets. *Journal of Economic Literature* 36 (4), 2021–2064.
- Shin, H. S., 1991. Optimal Betting Odds Against Insider Traders. *The Economic Journal* 191 (408), 1179–1185.
- Shin, H. S., 1992. Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias. *The Economic Journal* 102 (411), 426–435.
- Shin, H. S., 1993. Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims. *The Economic Journal* 103 (420), 1141–1153.
- Smith, M. A., Paton, D., Williams, L. V., 2006. Market Efficiency in Person-to-Person Betting. *Economica* 73 (292), 673–689.
- Snowberg, E., Wolfers, J., 2010. Explaining the Favorite–Long Shot Bias: Is it Risk-Love or Misperceptions? *Journal of Political Economy* 118 (4), 723–746.
- Tversky, A., Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5 (4), 297–323.
- Wolfers, J., Zitzewitz, E., 2004. Prediction Markets. *Journal of Economic Perspectives* 18 (2), 107–126.
- Woodland, L. M., Woodland, B. M., 1994. Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market. *The Journal of Finance* 49 (1), 269.
- Woodland, L. M., Woodland, B. M., 2003. The Reverse Favourite-longshot Bias and Market Efficiency in Major League Baseball: An Update. *Bulletin of Economic Research* 55 (2), 113–123.