ARTICLE IN PRESS
JID: EOR [m5G;April 28, 2018;9:29]
European Journal of Operational Research 000 (2018) 1–14

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing and Logistics

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, Southampton SO17 1BJ, UK
[b] Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK

ARTICLE INFO

Article history:
Received 18 April 2017
Accepted 6 April 2018
Available online xxx

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

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.

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., Arnesen & Bergfjord, 2014; Berg, Nelson, & Rietz, 2008).

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

☆ 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.
E-mail addresses: v.restocchi@soton.ac.uk (V. Restocchi), f.j.mcgroarty@soton.ac.uk (F. McGroarty), eg@ecs.soton.ac.uk (E. Gerding), j.e.johnson@soton.ac.uk (J.E.V. Johnson).

https://doi.org/10.1016/j.ejor.2018.04.011
0377-2217/© 2018 Published by Elsevier B.V.

Please cite this article as: V. Restocchi et al., It takes all sorts: A heterogeneous agent explanation for prediction market mispricing, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.04.011

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 behavioural 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 & Hommes, 1997; Buckley, Brown, & Marshall, 2012; Joëts, 2015; Lux & Marchesi, 1999).

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 behaviour of the crowd (Ma, Tang, McGroarty, Sung, & Johnson, 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, Tompkins, & Ziemba, 2003 for a study of the bias in S&P500 and FTSE100 index futures options, and Wolfers & 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 & 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 generalising 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, Olson, & Rappaport, 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 & Tamarkin, 1998) and misbelief (e.g. e.g. Gandhi & Serrano-Padial, 2015; Snowberg & Wolfers, 2010), 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, Gandhi, Salanié , and Salanié (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 & 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 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, Muller, and Schumacher (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

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