

# Does Belief Heterogeneity Explain Asset Prices: The Case of the Longshot Bias

AMIT GANDHI

*Department of Economics, University of Wisconsin-Madison*

and

RICARDO SERRANO-PADIAL

*First version received April 2011; final version accepted January 2014 (Eds.)*

This article studies belief heterogeneity in a benchmark competitive asset market: a market for Arrow–Debreu securities. We show that differences in agents’ beliefs lead to a systematic pricing pattern, the favourite–longshot bias (FLB): securities with a low-pay-out probability are overpriced, whereas securities with high probability pay-out are underpriced. We apply demand estimation techniques to betting market data, and find that the observed FLB is explained by a two-type population consisting of *canonical* traders, who hold virtually correct beliefs and are the majority type in the population (70%); and *noise* traders exhibiting significant belief dispersion. Furthermore, exploiting variation in public information across markets in our data set, we show that our belief heterogeneity model empirically outperforms existing preference-based explanations of the FLB.

*Key words:* Heterogeneity, Prospect theory, Favourite-longshot, Rational expectations, Demand estimation, Random utility, Noise traders, Risk preferences

*JEL Codes:* C13, C51, D40, G13, L00

## 1. INTRODUCTION

There are two important issues that have gained prominence in the study of financial markets. First, there is a growing consensus that the sheer trading volume we observe cannot be explained without resorting to information-driven trade (Cochrane, 2007), *i.e.* belief heterogeneity is likely a major source of gains from trade in financial markets. Second, there is also a growing body of evidence that documents “anomalies” in prices (see *e.g.* Keim, 2008), *i.e.* systematic patterns where prices do not accurately reflect the underlying fundamentals of securities.

In this article, we show these two issues are closely related: belief heterogeneity can serve as a natural foundation for pricing patterns that depart from the predictions of standard neoclassical theory. We focus on a particular pricing “anomaly”, known as the favourite–longshot bias (FLB), which has been found in a number of market settings, particularly in betting markets.<sup>1</sup> Betting markets are important for several reasons. Chief among them is that they represent real-world

1. There is evidence of the FLB in betting and prediction markets (see *e.g.* Griffith, 1949; Jullien and Salanié, 2000; Sauer, 1998; Snowberg and Wolfers, 2010, for a review), option markets (where the FLB is associated with the “volatility smile”), and in some derivative markets (Tompkins *et al.*, 2008).

Arrow–Debreu (A–D) security markets, which play a foundational role in the theoretical study of financial markets. In addition, unlike in other markets, we observe *ex post* security returns which, combined with the availability of large data sets, allows tracing prices to fundamentals.<sup>2</sup> There is a long empirical literature that widely documents the existence of the FLB in these markets—securities with a large probability of pay-out (favourites) yield higher average returns than securities with a small pay-out probability (longshots), with observed disparities in returns as high as 200%. Because of this, the FLB is often viewed as evidence against the standard notion of agents being *risk-averse, expected-utility* maximizers with *correct beliefs* (or rational expectations) about the underlying fundamentals (Thaler and Ziemba, 1988). This has led to two main alternative approaches to explain the observed pattern of returns, both of them based on departures from standard preferences while retaining rational expectations: one approach is to use convexity of utility, *i.e.* risk love (Quandt, 1986), whereas the other approach departs from expected utility theory altogether and uses non-linear probability weighting, *i.e.* rank-dependent utility or cumulative prospect theory (CPT) preferences (Jullien and Salanié, 2000; Snowberg and Wolfers, 2010; Thaler and Ziemba, 1988).

In contrast to these preference-based approaches, this article studies the behaviour of asset prices when agents have heterogeneous beliefs. In particular, we use the standard A–D security market setting and explore whether belief heterogeneity, apart from rationalizing trade, can also explain the existence of the FLB. We then investigate whether our belief heterogeneity explanation can be empirically distinguished from existing preference-based explanations of the FLB using data from U.S. racetrack betting. Our main findings can be summarized as follows.

First, we theoretically show that belief heterogeneity among risk neutral traders with finite wealth naturally leads to the FLB in competitive A–D security markets: as long as the correct beliefs about the true pay-out probabilities lie in the support of the population belief distribution, favourites will be underpriced and longshots overpriced.

Secondly, we present a framework for estimating belief heterogeneity in A–D markets using aggregate data on prices and *ex post* returns. In particular, we show that our theoretical model of belief heterogeneity can be expressed as a random utility model that we are able to non-parametrically identify from variation in prices and *ex post* returns in the data. We then provide an estimation strategy by assuming that the random utility component follows a variance mixture of logit models. The variance mixture allows us to identify different types of agents in the underlying population as sub-populations with different dispersion in beliefs. Using betting market data from horse racing in the U.S., we find that a two-type population of risk-neutral traders, in which the prevalent type (about 70% of agents) holds virtually correct beliefs with minimal belief dispersion and the other type exhibits higher belief dispersion, explains the observed pricing pattern remarkably well. This provides suggestive evidence of the existence of two types of agents in these markets that trade in equilibrium—perfectly and imperfectly informed or *noise* traders.

Finally, given our empirical findings, we compare our belief heterogeneity hypothesis to existing preference-based explanations for the FLB. We do so first by parametrically comparing our two-type belief model with a representative agent whose preferences follow CPT, which is a leading preference explanation in the literature. In particular, we use the same specification of CPT as the empirical analysis of Jullien and Salanié (2000), which has the same number of parameters as our two-type belief model and thus serves as a convenient baseline for comparison. We fail to reject the hypothesis that the predicted probabilities emerging from our model equal

2. An additional key feature of betting markets is the absence of differences in liquidity or transaction cost across securities. This is in contrast with other markets where pricing patterns are closely associated with differences in liquidity and trading costs (see *e.g.* Sadka and Scherbina, 2007, and references therein).

the true non-parametric probabilities in the data, but find a significant rejection for CPT. Model comparisons using Vuong non-nested hypotheses tests also support the better explanatory power of the beliefs model.

We then aim at providing a more general contrast between our approach and existing preference-based explanations (which may include heterogeneity of preferences). The fundamental distinction between beliefs and preferences as a driver of choice is that beliefs should respond to information, whereas preferences should not. We exploit this distinction by using a source of variation in the amount of information at races. Specifically, races that are run at the same track on the same day take two different forms, maiden and non-maiden. Because only relatively experienced horses can participate in non-maiden races and new horses (without any racing history) participate in maiden races, traders in non-maiden races have access to richer information about the value of the A–D securities associated with each horse. We find that this difference in the information structure across race types is starkly reflected in prices: maiden races exhibit a much more pronounced FLB as compared to non-maiden races. We estimate our model on each race sub-sample and show that it explains this pricing difference in the natural way: the same proportions of informed and noise traders are present in both types of races, but beliefs are more dispersed in the low-information races. Preference-based theories that rely on homogeneous correct beliefs, on the other hand, cannot easily account for the price differences across information environments: they would require a dramatic change in agent preferences across races, which we also show cannot be reconciled by self-selection into races.

What is the intuition for why belief heterogeneity explains the FLB? Belief heterogeneity can generate a dramatic disparity in returns even when beliefs are heavily concentrated around correct beliefs because the FLB, like many pricing puzzles observed in financial markets, is driven by behaviour at the tails of the underlying value distribution: disparity in expected returns becomes most apparent for securities with pay-out probabilities  $< 1\%$ . Accordingly, a little dispersion in beliefs suffices to induce enough demand on those longshots to generate substantial overpricing in equilibrium. For example, in our data, the average security has over twice as high an expected return as compared to extreme longshots that pay-out  $< 1\%$  of the time. In our estimated model, only roughly 5% of agents would prefer these longshots over the average security. But this small demand is sufficient to generate the observed overpricing.

Our findings highlight the potentially misleading inferences that can be drawn from representative agent models. Although a representative agent may rationalize the aggregate demand of a population of heterogeneous agents, his/her preferences need not reflect those in the population in any meaningful sense. For example, in our model, the modal behaviour is well captured by the textbook rational expectations, risk-neutral agent. In contrast, a representative agent exhibits non-standard preferences or beliefs to explain the same data.

**Related Literature:** Our theoretical model is related to existing models linking belief heterogeneity to the FLB, notably the competitive model of [Ali \(1977\)](#) and the asymmetric information approach of [Ottaviani and Sorensen \(2006, 2010\)](#), but also the work of [Shin \(1991, 1992\)](#) and [Potters and Wit \(1996\)](#). The main difference between our approach and the existing literature is our focus on behaviour at the tails of the value distribution, where the FLB empirically arises. In contrast, existing research typically defines the FLB for the whole range of values, *e.g.* by predicting returns to be strictly increasing in pay-out probabilities. Accordingly, we are able to substantially weaken the conditions under which the FLB arises. In addition, unlike [Ottaviani and Sorensen \(2006, 2010\)](#), our goal is not to provide an informational foundation to heterogeneous posterior beliefs in asset markets, but rather find general conditions on the distribution of such beliefs that lead to the FLB.

Our work is also related to few papers showing that asset prices in a population of agents with heterogeneous beliefs need not “represent” the underlying distribution of beliefs. In particular, [Manski \(2006\)](#) shows that in a competitive, binary A–D security market, the equilibrium price does not reflect the average opinion in the population. In the same setting, [Ottaviani and Sorensen \(2012\)](#) show that market prices systematically under-react to new information relative to the Bayesian agents in the population. Our results evoke a similar message in that we highlight the potentially misleading inferences that can be drawn from representative agent models. Nevertheless, an important difference between our article and these stems from the fact that our model links the distribution of beliefs to the underlying fundamental probabilities, which is needed to explain the FLB. By contrast, these papers do not model the relationship between beliefs and fundamentals and focus on the aggregation of opinions rather than on the FLB.

Importantly, this article adds a novel dimension to the empirical literature on the FLB, which, to date, has emphasized preference-based explanations ([Golec and Tamarkin, 1998](#); [Jullien and Salanié, 2000](#); [Snowberg and Wolfers, 2010](#)). Methodologically, our estimation strategy builds on the work of [Jullien and Salanié \(2000\)](#), who propose a maximum likelihood framework to estimate representative agent preferences using data on prices and *ex post* returns. We add heterogeneity to this framework and show that the resulting random utility model can be non-parametrically identified. A key feature of our approach is that randomness arises from belief heterogeneity, and belief-based trade has not yet been explored as an empirical alternative. Relatedly, our finding that the magnitude of the FLB changes across races characterized by different information structures appears to be a new contribution to the literature on betting and financial markets.

The recent work by [Chiappori et al. \(2009, 2012\)](#) also add heterogeneity to the framework of [Jullien and Salanié \(2000\)](#) but from the perspective of preferences rather than beliefs. There are at least two important differences between their paper and our beliefs approach. First, while the discrete choice assumption on bettor behaviour is without loss of generality in our model because we use risk-neutral agents, it is imposed as a restriction in their set-up, effectively preventing risk-averse agents from constructing portfolios of bets to reduce their exposure to risk. Secondly, our empirical model of belief heterogeneity is isomorphic to a model of horizontally differentiated demand (with a margin of indifference between any two assets), whereas their empirical model of preference heterogeneity in risk tastes is isomorphic to a vertically differentiated demand model (with a margin of indifference only between two assets with adjacent prices). This gives rise to rather different identification and estimation strategies. Nevertheless, they also find that standard risk preferences cannot easily explain the U.S. racetrack-betting data even after accounting for preference heterogeneity, which is thus complementary to our findings.

The plan of the article is as follows. In Section 2, we describe the market and characterize the demand of a risk-neutral agent. In Section 3, we introduce belief heterogeneity and show that it produces the FLB in equilibrium. In Section 4, we describe the data set and illustrate the observed price pattern. We show in Section 5 how to move from the theory to an empirical framework for measuring heterogeneity. In Section 6, we describe our estimation results, while we compare our approach to the leading preference-based alternatives in Section 7. Section 8 concludes.

## 2. THE MARKET

We consider a competitive market for  $n$  A–D securities. Security  $i$  pays \$1 if outcome  $i$  takes place. Before the market opens, nature determines the state  $\mathbf{p} = (p_1, \dots, p_n) \in \text{int } \Delta^{n-1}$ , where  $\text{int } \Delta^{n-1}$  is the interior of the  $(n-1)$  dimensional simplex and  $p_i$  is the probability that outcome  $i$  is realized. That is,  $\Delta^{n-1}$  is the set of all possible probability distributions and the state  $\mathbf{p} \gg 0$  is a particular probability distribution over the  $n$  possible outcomes associated with the A–D securities.

Let  $\rho_i$  denote the price of security  $i$ . We allow for the possibility of positive transaction costs represented by a fraction  $\tau \geq 0$  of each dollar invested in the market that the institution keeps for its own profit. In this context, the expected gross return of investing \$1 in security  $i = 1, \dots, n$  is given by

$$ER_i = (1 - \tau) \frac{P_i}{\rho_i}, \quad (1)$$

and the expected net return is  $ER_i - 1$ . Accordingly, securities that are underpriced relative to their true chance of yielding returns have a higher expected return as compared to securities that are overpriced.

The market consists of a population  $T = [0, 1]$  with a continuum of risk-neutral agents with finite endowments. Assume for the moment that there are no transaction costs ( $\tau = 0$ ) and that the outside option yields zero net returns given agent's beliefs. The demand of agent  $t \in T$  is a bundle  $(x_1, \dots, x_n)$  such that  $\sum x_i \leq w_t$ , where  $x_i \geq 0$  is the amount invested on security  $i$ , and  $w_t > 0$  is the dollar endowment of the agent.

Due to risk neutrality, the demand for security  $i$  of any agent  $t$  with well-defined posterior beliefs, denoted by  $(\pi_{1t}, \dots, \pi_{nt})$ , is determined by the relative comparisons of *subjective* expected returns, given by  $\pi_{it}/\rho_i$ ,  $i = 1, \dots, n$ . That is, agent  $t$  will invest on the security  $i$  if it yields the highest subjective return, *i.e.* if  $\pi_{it}/\rho_i > \pi_{jt}/\rho_j$  for all  $j \neq i$ .

We assume risk neutrality for two reasons. First, unlike risk aversion, it rationalizes the discrete choice of investing all the endowment in one of the securities, rather than hedging or investing only a fraction of it. This is important given that most empirical studies of the FLB take a discrete-choice approach due to the fact that endowments are not observed (see for instance [Jullien and Salanié \(2000\)](#), [Snowberg and Wolfers, 2010](#)). Secondly, people tend to be risk-neutral when stakes are low (*e.g.* [Bombardini and Trebbi, 2010](#)), as it is often the case for most traders in betting markets. Nonetheless, we allow for alternative risk attitudes in our estimation and find that, while introducing a risk aversion parameter improves the empirical fit, the absolute risk aversion coefficient is virtually zero in the presence of belief heterogeneity (see Section 6).

Let  $r_{it} = p_i/\pi_{it}$ , which reflects agent  $t$ 's relative deviation from correct beliefs about security  $i$ . The next lemma shows that the demand of agent  $t$  for security  $i$  can be characterized instead in terms of the *objective* expected returns and a vector of *indifference ratios*, given by  $r_{it}/r_{jt}$  for  $j \neq i$ : the agent is indifferent between security  $i$  and  $j$  whenever the ratio of expected returns  $ER_i/ER_j$  is equal to the indifference ratio  $r_{it}/r_{jt}$ .

**Lemma 1.** *There exists a vector  $\mathbf{r}_t = (r_{1t}, \dots, r_{nt}) > 0$  with  $r_{it} = p_i/\pi_{it}$  such that, for all  $(\rho_1, \dots, \rho_n) \gg 0$  and all  $i = 1, \dots, n$ , the demand of a risk-neutral agent  $t$  satisfies*

$$q_{it}(ER_1, \dots, ER_n) = \begin{cases} w_t & \frac{ER_i}{ER_j} > \frac{r_{it}}{r_{jt}} \forall j \neq i \\ \{x : x \leq w_t\} & \frac{ER_i}{ER_j} \geq \frac{r_{it}}{r_{jt}} \forall j, \frac{ER_i}{ER_j} = \frac{r_{it}}{r_{jt}} \text{ some } j \neq i \\ 0 & \frac{ER_i}{ER_j} < \frac{r_{it}}{r_{jt}} \text{ some } j \neq i, \end{cases}$$

and

$$\sum_i q_{it}(ER_1, \dots, ER_n) \leq w_t,$$

which holds with equality if  $\frac{ER_i}{ER_j} \neq \frac{r_{it}}{r_{jt}}$  for some  $i, j$ .

*Proof* In the Appendix.  $\parallel$

Notice that an agent with correct beliefs has indifference ratios equal to 1 for all  $i, j$ . Likewise, if an agent has an indifference ratio  $r_{1t}/r_{2t} = 2$ , then asset 1 would have to pay twice as high objective expected returns as asset 2 for the agent to be indifferent between the two (thus the agent either seriously overestimates the probability of asset two paying off, or underestimates asset 1).

There are two advantages of working with objective returns and indifference ratios rather than dealing directly with prices and subjective beliefs. First, the fact that indifference ratios reflect relative differences with respect to the true underlying pay-out probabilities allows us to define (idiosyncratic) belief heterogeneity in terms of deviations from correct beliefs (see Assumption 1 below). Secondly, the indifference conditions in Lemma 1 readily lead to the random utility model we use later in our empirical analysis.

Finally, let  $s_i$  be security  $i$ 's market share, *i.e.* the amount invested on security  $i$  relative to the total (finite) amount invested in the market. Market clearing implies that prices equal market shares, as long as agents, when indifferent between investing or not, choose to invest either all their endowment or nothing.<sup>3</sup> It is worth noting that, by definition, prices are equal to shares in parimutuel markets, as are the markets comprising our data set.

### 3. BELIEF HETEROGENEITY AND PRICES

In a standard representative agent model, traders are homogeneous in terms of beliefs and preferences. In particular, they hold correct posterior beliefs about the state of the world. We call such (risk-neutral) agents *canonical traders*, and their individual demand is characterized by  $\mathbf{r}_t = (1, \dots, 1)$ . In this context, the only trading equilibrium outcome is  $ER_i = ER_j$  for all  $i, j$ , since otherwise the market does not clear.<sup>4</sup> That is, there do not exist any gains from trade in this economy and agents are indifferent between investing in any of the securities or staying out of the market, and thus such equilibrium is not robust to the introduction of transaction costs.

Now consider introducing gains from trade in this baseline model by letting agents exhibit differences in posterior beliefs, *i.e.* heterogeneity in indifference ratios. Let the mass function of  $\mathbf{r}_t$  be given by the conditional probability measure  $P[\cdot | \mathbf{p}, \theta]$ , which depends both on the state of the world  $p$  and possibly on other characteristics of the market  $\theta \in \Theta$ . We only require that there is always a positive mass to “both” sides of the canonical trader—this avoids the no-trade trap discussed above. To capture this condition, given  $z \in \mathbb{R}_+^{n-1}$ , let

$$L_i[z | \mathbf{p}, \theta] = P \left[ \left( \frac{r_{it}}{r_{1t}}, \dots, \frac{r_{it}}{r_{i-1t}}, \frac{r_{it}}{r_{i+1t}}, \dots, \frac{r_{it}}{r_{nt}} \right) \ll z | \mathbf{p}, \theta \right], \text{ for } i = 2, \dots, n-1,$$

and define  $L_1$  and  $L_n$  in a similar fashion.  $L_i[z | \mathbf{p}, \theta]$  represents the mass of agents with indifference ratios associated with security  $i$  that are lower than  $z$ , *i.e.*  $r_i/r_j < z_j$  for  $j \neq i$ . For instance,

3. Market clearing means the supply of dollars equals the demand of dollars in each of the possible  $n$  outcomes. This happens if and only if

$$\frac{q_i}{\rho_i} = (q_1 + \dots + q_n) \iff \rho_i = s_i = \frac{q_i}{q_1 + \dots + q_n} \quad (\forall i),$$

where  $q_i$  represents the aggregate investment in dollars on security  $i$ . If a mass of agents is indifferent between investing in some securities and staying out of the market, and they invest only a fraction of their endowment, then the supply of dollars may be higher than the total investment in the market, in which case some agents are short-selling some of the securities.

4. If  $ER_i > \max_{j \neq i} ER_j$  for some  $i$  then all traders invest in security  $i$ , leading to  $\rho_i = 1$  and  $\rho_j = 0$  for all  $j \neq i$ , implying  $ER_j > ER_i$ , a contradiction.

$L_i[(1, \dots, 1) | \mathbf{p}, \theta]$  is the mass of agents that strictly prefer  $i$  over any other security when expected returns are equal across securities. In this context, we assume that there is enough heterogeneity so that demand for security  $i$  is bounded away from 0 and 1 when  $\tau = 0$  and  $ER_i = 1$  for all  $i = 1, \dots, n$ . That is, there is minimal “liquidity” at fair prices regardless of the market characteristics  $(\mathbf{p}, \theta)$ .

**Assumption 1.** (Idiosyncratic belief heterogeneity)  $L_i[(1, \dots, 1) | \cdot]$  is bounded away from zero for  $i = 1, \dots, n$ .

It is worth emphasizing that this assumption allows for the support of  $L_i$  to converge to a single point as  $p_i \rightarrow 0$ , as long as all the mass is not concentrated at  $(1, \dots, 1)$  when  $p_i > 0$ . For instance,  $L_i$  could have full support in  $[1 - \varepsilon, 1 + \varepsilon]^{n-1}$  with  $\varepsilon \rightarrow 0$  as  $p_i \rightarrow 0$ . In other words, it allows for subjective beliefs to be absolutely continuous with respect to true probabilities, and it does not implicitly require the distribution of subjective beliefs to exhibit “fat tails”. In addition, as we show in the Appendix, Assumption 1 can be weakened when endowments and beliefs are independently distributed, by letting the distribution of beliefs exhibit “vanishing tails”, and our results would still go through.

The next result shows that heterogeneity induces the FLB, even in the presence of transaction costs. In particular, any security  $i$  with a sufficiently high value ( $p_i$ ) is underpriced in equilibrium, while the remaining (low-value) securities are overpriced on average.

**Theorem 1.** (FLB) *If Assumption 1 holds, there exists  $\bar{\tau} > 0$  such that for all  $\tau < \bar{\tau}$  a necessary consequence of equilibrium is that there exists  $\bar{q} < 1$  such that, for all  $i = 1, \dots, n$ , if  $p_i > \bar{q}$  then security  $i$  is underpriced, while securities  $j \neq i$  are overpriced on average.*

*Proof* In the Appendix ||

Why does the FLB arise when agents are heterogeneous? Three key features in our model help explain it. On the institutional side, the underlying value of an A–D security is bounded, since it is given by its pay-out probability. On the agent side, we assume that agents have finite endowments and are risk neutral. Hence, they focus on relative comparisons of (subjective) expected returns across securities, choosing to invest their endowment on the highest return security. Finally, we assume that belief heterogeneity is idiosyncratic, *i.e.* represented by some dispersion of subjective expected returns around the true underlying expected returns. In this context, risk neutrality ensures that demand for each security does not vanish when true expected returns are equal across securities, regardless of pay-out probabilities. As a consequence, the presence of this minimal demand pushes the price of a security above its pay-out probability whenever the latter is very small, causing the security to be overpriced—the underpricing of securities with a high pay-out probability directly follows by noting that prices in the market add up to 1.

We provide some intuition by focusing on the case of two securities, unit endowments ( $w_t = 1 \forall t \in T = [0, 1]$ ), and zero transaction costs. Note that, in this context, if the returns across the securities are  $(ER_1, ER_2)$ , then the share  $s_1$  of investment in security 1 is bounded above and below by

$$L_1[ER_1/ER_2 | \mathbf{p}, \theta] \leq s_1 \leq 1 - L_2[ER_2/ER_1 | \mathbf{p}, \theta].$$

These bounds are tight because agents with  $r_{1t}/r_{2t} = ER_1/ER_2$  are indifferent towards any investment  $(x_1, x_2)$  with  $0 \leq x_1 + x_2 \leq 1$ .

In this market, idiosyncratic heterogeneity implies the existence of a positive lower bound of demand for security 1 when  $ER_1 = ER_2$ , given by  $\underline{q} = \inf \{L_1[1 | \mathbf{p}, \theta] : (\mathbf{p}, \theta) \in \text{int } \Delta \times \Theta\}$ . But then, when  $p_1$  is sufficiently low, this lower bound  $\underline{q}$  on demand prevents the share of security  $i$  and thus

its price to fall below  $p_1$ , leading to overpricing of security 1 ( $ER_1 < 1$ ) and underpricing of security 2 ( $ER_2 > 1$  since  $p_2 = 1 - p_1 < 1 - s_1 = s_2$ ). More formally, let  $p_1 < q$  and suppose the theorem does not hold, *i.e.*  $ER_1 \geq ER_2$ . Then, by the above bounds we must have  $s_1 \geq L_1[1 | \mathbf{p}, \theta] \geq q > p_1$ . But this is a contradiction since  $s_1 > p_1$  implies that  $ER_1 < ER_2$ .<sup>5</sup> A similar intuition applies when  $p_1$  is sufficiently high, given the upper bound  $1 - L_2$ .

It is worthwhile to note the generality of our condition. Idiosyncratic heterogeneity only requires that there is some dispersion of beliefs around the correct beliefs, *i.e.* that there are minimal gains from trade in each security at fair prices.

To illustrate the potential of belief heterogeneity to generate the FLB consider the following example. Suppose there are two assets in the market with probability of paying out a dollar being 0.9% and 99.1%, respectively, and there are 99 risk-neutral agents with accurate beliefs and only one risk-neutral agent who believes that the probability of pay-out on the longshot is anything better than 1%. If all agents have unit endowments, then the trading equilibrium involves this single agent investing in the longshot while all the remaining agents invest in the favourite. In this case, the A–D price is 0.01 for asset 1 and 0.99 for asset 2. Thus, a very small departure in beliefs of 1% of the population creates a sizeable disparity in returns: the expected return on the longshot in the heterogeneous agent economy is  $-10\%$ , whereas the favourite has a positive expected return.

Theorem 1 states that the FLB arises for extreme probabilities. However, it would be interesting to know under what conditions a global FLB arises, *i.e.* that securities with higher pay-out probabilities exhibit higher equilibrium returns. As the next result shows, one way to obtain the global FLB is to impose a symmetry condition on the distribution of beliefs. Specifically, to require that, whenever asset  $i$  equilibrium returns are at least as high as asset  $j$ 's, the mass of agents believing that asset  $i$  yields the highest returns is at least as large as those believing the best asset is  $j$ .

**Assumption 2.**  $P[\cdot | \mathbf{p}, \theta]$  is atomless for all  $(\mathbf{p}, \theta)$ . Moreover, if  $ER_i \geq ER_j$  then

$$L_i \left[ \frac{ER_i}{ER_1}, \dots, \frac{ER_i}{ER_n} | \mathbf{p}, \theta \right] \geq L_j \left[ \frac{ER_j}{ER_1}, \dots, \frac{ER_j}{ER_n} | \mathbf{p}, \theta \right]$$

for all  $i, j = 2, \dots, n$ , with a similar condition for  $i, j \in \{1, n\}$ .

**Theorem 2.** (Global FLB) If Assumption 2 holds, there exists  $\bar{\tau} > 0$  such that, for all  $\tau < \bar{\tau}$ ,  $p_i < p_j$  implies  $ER_i < ER_j$  in equilibrium.

*Proof* We only proof the case of  $\tau = 0$ . The proof for  $\tau < \bar{\tau}$  follows the same argument used in the proof of Theorem 1 and is therefore omitted.

First, notice that  $P$  being atomless implies that if equilibrium expected returns are given by  $ER_i$  for all  $i$  then market shares satisfy  $s_i = L_i \left[ \frac{ER_i}{ER_1}, \dots, \frac{ER_i}{ER_n} | \mathbf{p}, \theta \right]$ . Now consider, by way of contradiction, that  $p_i < p_j$  but  $ER_i \geq ER_j$ . Then it must be the case that  $s_i < s_j$  since  $ER_i = p_i/s_i \geq p_j/s_j = ER_j$ . But this implies that

$$L_i \left[ \frac{ER_i}{ER_1}, \dots, \frac{ER_i}{ER_n} | \mathbf{p}, \theta \right] < L_j \left[ \frac{ER_j}{ER_1}, \dots, \frac{ER_j}{ER_n} | \mathbf{p}, \theta \right],$$

a violation of Assumption 2.  $\parallel$

5. Notice that all that we need is  $L_1[1 | \mathbf{p}, \theta] > \alpha p_1$  for some  $\alpha > 1$  and all  $p_1$  small enough. Hence, we can replace the uniform bound on  $L_1$  with this “vanishing bound”, as long as  $s_1 \geq L_1$  holds, which is the case when endowments and beliefs are independent. See Appendix A.1 for details.



This result is a generalization of Ali (1977) to the case of  $n \geq 2$  horses. He shows that the global FLB arises in a market with two securities whenever  $Pr(r_{1t}/r_{2t} \leq 1 | \mathbf{p}, \theta) = 0.5$  for any  $(\mathbf{p}, \theta)$ , *i.e.* the canonical belief is the median belief in the population. To see why Theorem 2 generalizes this result notice that, when  $n = 2$ , by Assumption 2 we must have that  $L_1[1 | \mathbf{p}, \theta] = L_2[1 | \mathbf{p}, \theta] = 1 - L_1[1 | \mathbf{p}, \theta]$ , *i.e.*  $L_1[1 | \mathbf{p}, \theta] = 1/2$ . Using a similar argument it is easy to show that Assumption 2 implies Assumption 1: at fair prices we must have that  $L_i[(1, \dots, 1) | \mathbf{p}, \theta] = L_j[(1, \dots, 1) | \mathbf{p}, \theta]$  for all  $i, j$ . But since  $P$  is atomless then  $L_i[(1, \dots, 1) | \mathbf{p}, \theta] = 1/n$  for all  $i$  and all  $(\mathbf{p}, \theta)$ .

#### 4. DATA

Betting markets, and racetrack betting in particular, are textbook illustrations of A–D securities markets. We focus our attention on the “win odds” market, which is the market for bets on which horse will win. This is considered the most competitive market at the racetrack, given it has the most liquid pool of money. Not surprisingly, it has been the subject of most of the empirical attention.

Prices at the racetrack are quoted in terms of the odds  $R_i$  on horse or security  $i = 1, \dots, n$ , which are defined as the *net* return per dollar bet on security  $i$  in the event  $i$  wins the race—the *gross* return is given by  $R_i + 1$ . In North American racetracks, the odds are determined through a “parimutuel” system of wagering in which the losers pay the winners. This system ensures that there are no differences in liquidity across securities and is equivalent to the market clearing condition in an A–D securities market where shares equal prices. Accordingly, market odds satisfy

$$R_i = \frac{1 - \tau}{s_i} - 1 \quad i = 1, \dots, n. \quad (2)$$

Given a vector of observed odds  $(R_1, \dots, R_n)$ , we can invert equation (2) to recover the underlying A–D prices  $\rho_i = s_i$  for  $i = 1, \dots, n$ .

The FLB is in fact a widely documented empirical pattern of returns across bets at the racetrack. To illustrate it, consider a large data set consisting of horse starts, *i.e.* a sample of horses that competed in some race. Each observation  $i$  corresponds to a pair  $(win_i, R_i + 1)$ , where  $R_i + 1$  is the gross return that A–D security  $i$  pays conditional on horse  $i$  winning and  $win_i$  is an indicator variable for whether horse  $i$  won the race or not. Define  $A_i$  as the *ex post* gross return on security  $i$ , *i.e.*

$$A_i = \begin{cases} R_i + 1 & \text{if } win_i = 1 \\ 0 & \text{if } win_i = 0, \end{cases}$$

and thus the regression

$$E[A_i | \log s_i] = E[(R_i + 1) win_i | \log s_i] = f(\log s_i) \quad (3)$$

measures the expected gross return among securities that have the same log price ( $\log s_i$ ). Using a sample of 176,652 races that were collected from North American tracks over 2003–2006,<sup>6</sup> consisting of 1,456,512 horse starts, we estimate  $f$  by non-parametrically regressing  $A_i$  on  $\log s_i$  using a locally linear kernel weighted regression.<sup>7</sup> Figure 1 shows the estimated returns—along

6. The data were collected through the efforts of members at the website [paceadvantage.com](http://paceadvantage.com). This data set is one of the largest to have been assembled to study betting behaviour.

7. We use the Epanechnikov kernel and a rule of thumb bandwidth estimator.

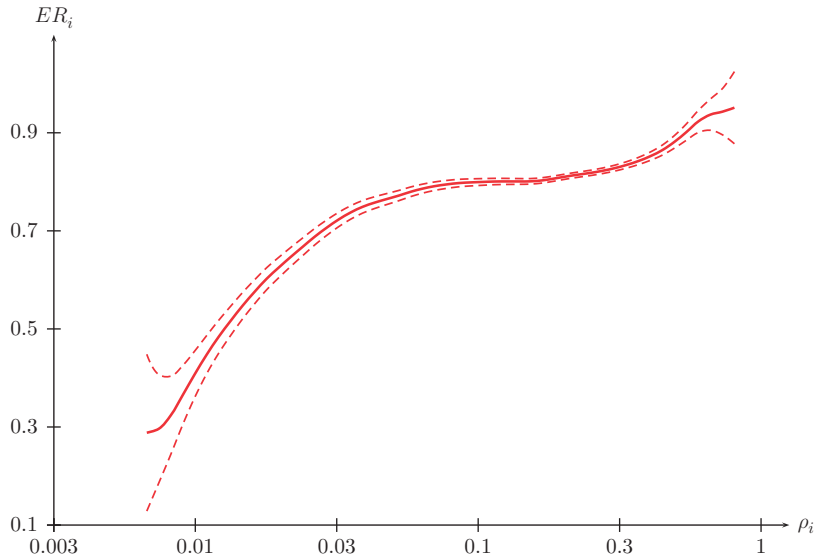


FIGURE 1  
Favourite longshot bias

with their 95% confidence interval, which clearly exhibit the FLB: low-value securities have lower expected returns, where the effect is particularly pronounced when comparing the extremes. The returns in Figure 1 are net of the track take. Thus, since the average track take is  $\sim 0.19$ , these returns imply that, roughly, horses with prices  $< 0.2$  are overpriced and those with prices  $> 0.2$  are underpriced.

To understand the argument implicitly put forth by [Thaler and Ziemba \(1988\)](#) as to why the FLB as found in the data is so puzzling, recall Lemma 1 and consider the indifference ratios of a representative agent that clears the market in equilibrium. In equilibrium, the agent must be indifferent across securities to sustain positive asset demand. The expected gross return to an average bet in the data is  $ER_{avg} = 0.76$ . We can thus use Figure 1 to recover the representative agent's indifference ratio ( $\frac{r}{r_\rho}$ ), which is the ratio associated with the "average" security (representing a random bet) and a security with price  $\rho$ . This ratio represents the amount that the average security must pay in expectation relative to a security with price  $\rho$  in order for the agent to be indifferent, and we can back out this relationship from the data because the representative agent is indifferent across all securities for the market to clear. As Figure 2 shows, the representative agent overvalues securities with prices  $\leq 0.01$  by  $> 50\%$ , and securities with prices  $\leq 0.005$  by  $\sim 100\%$ , compared to the average bet.

## 5. ESTIMATING HETEROGENEITY

We now show how to estimate a model of an A–D economy where agents have heterogeneous beliefs. The goal is to estimate the pattern of belief heterogeneity that is consistent with both the betting data and equilibrium in the model. That is, we seek to understand whether a heterogeneous beliefs model can explain the observed FLB and, furthermore, what the estimated heterogeneity looks like in contrast with the above representative agent.

To illustrate our empirical strategy, we first describe the general maximum likelihood (ML) approach for estimating preferences from aggregate betting data, which was first introduced by

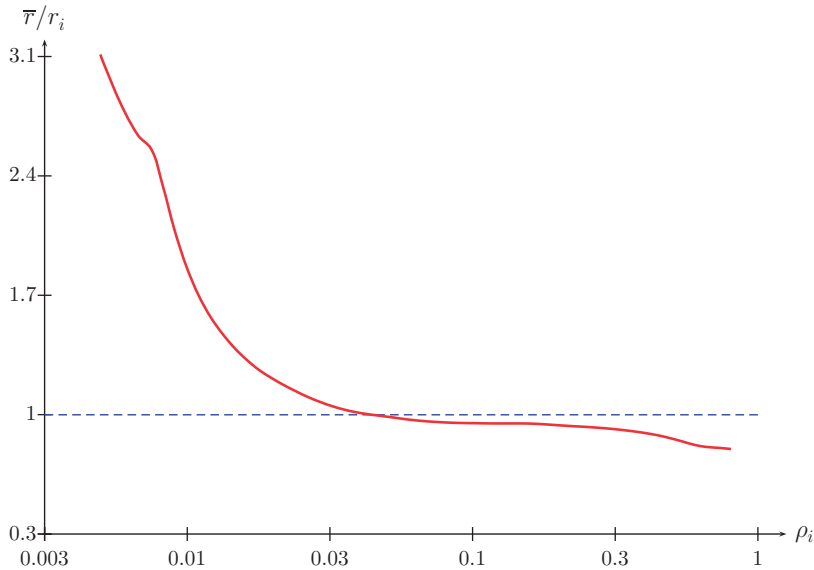


FIGURE 2  
Representative agent indifference ratios

Jullien and Salanié (2000) (henceforth referred to as JS) in the context of a representative agent framework. We then show how to extend their framework to a population of heterogeneous agents. Let us recall that the data-set consists of a sample of  $K$  markets or races, which are assumed to be independent of each other. Each market  $k = 1, \dots, K$  is defined by the number of securities  $n^k$ , a vector of odds  $\mathbf{R}^k = (R_1^k, \dots, R_{n^k}^k)$ , and the identity of the winner  $i_{win}^k$ . The data thus identify the empirical relationship between prices and fundamentals, *i.e.* for any number of securities  $n$  with odds  $(R_1, \dots, R_n)$  we can identify the underlying probabilities of winning  $\mathbf{p}(R_1, \dots, R_n)$ : intuitively,  $p_i(R_1, \dots, R_n)$  is identified by the fraction of times horse  $i$  wins in the subset of races characterized by  $(R_1, \dots, R_n)$ . Formally, this can be identified by the non-parametric regression of the indicator variable that a horse  $i$  wins on the vector of odds in the race.

A key insight in JS is to recognize that an equilibrium model of the betting market, described by parameters  $\theta \in \Theta$ , implies a relationship between prices and fundamentals, denoted  $\mathbf{p}(R_1, \dots, R_n; \theta)$ . The empirical strategy implicitly underlying JS is to find the parameters  $\theta \in \Theta$  whose prediction about the equilibrium relationship between prices and fundamentals best matches the actual empirical relationship. This strategy is naturally implemented using maximum likelihood, and can be described by two key steps:

1. Given a choice of model parameters  $\theta \in \Theta$ , for each market  $k$ , find the vector of *unobserved* pay-out probabilities  $\mathbf{p}^k(\theta) = (p_1^k(\theta), \dots, p_{n^k}^k(\theta))$  that is consistent with *observed* odds  $\mathbf{R}^k$  under equilibrium. That is, we solve

$$\mathbf{p}^k(\theta) = \phi(\mathbf{R}^k, \theta), \quad k = 1, \dots, K, \quad (4)$$

where  $\phi$  is a model-specific mapping from odds to probabilities.

2. Estimate  $\theta$  by maximizing the log-likelihood

$$LL(\theta) = \sum_{k=1}^K \log p_{i_{win}^k}^k(\theta). \quad (5)$$

This ML estimator consistently estimates the value of  $\theta_0 \in \Theta$  that minimizes the Kullback–Leibler distance between the model  $\mathbf{p}(R_1, \dots, R_n; \theta_0)$  and the data  $\mathbf{p}(R_1, \dots, R_n)$ .<sup>8</sup>

JS showed that, in a representative agent framework of market equilibrium, there exists a unique equilibrium mapping between prices and fundamentals that is required for Step 1. In particular, the equilibrium mapping is determined by the set of indifference conditions

$$U(p_1^k, R_1^k; \theta) = U(p_2^k, R_2^k; \theta) = \dots = U(p_j^k, R_j^k; \theta), \quad (6)$$

where  $U(p_i^k, R_i^k; \theta)$  is the pay off to the agent from investing in a security with pay-out probability  $p_i^k$  at odds  $R_i^k$ . That is,  $\mathbf{p}^k$  is solved to make the agent indifferent about investing across securities at the observed odds. JS formally show that, if  $U$  is continuous and respects first-order stochastic dominance, the mapping  $\phi$  from odds to probabilities implied by (6) is well defined. Using these conditions, preference parameters  $\theta$ , such as risk aversion or CPT coefficients, can be estimated. Using this approach, JS compared different representative agent preferences, and found evidence in favour of CPT over risk loving as the preferred explanation of the FLB.

A heterogeneous agents model, however, cannot be approached using this representative agent framework: the set of agents who are indifferent across all securities in a heterogeneous population typically has measure zero, as is the case when the distribution of beliefs is continuous. Thus, we can no longer use the equilibrium conditions (6) to solve for the relationship between prices and probabilities  $\mathbf{p}^k(\theta) = \phi(\mathbf{R}^k, \theta)$  in a heterogeneous agents model (where the parameter  $\theta \in \Theta$  determines the pattern of heterogeneity in the underlying betting population). Instead, the additional information that we will use to solve for the equilibrium relationship between prices and fundamentals are the market shares of the securities  $s^k = (s_1^k, \dots, s_n^k)$ , which are derived from observed odds using equation (2). Note that market shares are not needed in the representative agent approach. However a heterogeneous agents model predicts market shares in a natural way.

In particular, if  $U_t$  denotes the pay off function of agent  $t$  and we assume that she invests his/her endowment on a single security,<sup>9</sup> market shares are given by the aggregation of individual investments:

$$s_i^k = \frac{1}{\int_T w_t dt} \int_T w_t \mathbf{1} \left[ U_t(p_i^k, R_i^k) > U_t(p_j^k, R_j^k) \forall j \neq i \right] dt, \quad i = 1, \dots, n^k, \quad (7)$$

where  $\mathbf{1}[\cdot]$  is the indicator function. Thus, our goal is to characterize  $U_t$  in terms of our model of belief heterogeneity and show that the system of equations (7) can be uniquely inverted to recover  $\mathbf{p}^k$ . To be consistent with our theoretical model and to ease exposition, we focus here on the case of a population of risk-neutral agents. Nonetheless, we describe in the Appendix

8. Standard errors follow from standard likelihood theory, which JS give explicitly for this context in Section 4.c of their paper.

9. While our theoretical analysis allows agents to abstain from the market when  $\tau > 0$ , here we assume that participation in the market is exogenously given and focus only on the decision of which asset to invest on. This is primarily driven by the data: we do not have information on the fraction of consumers who do not participate in the market. Rather, our data only give us demand conditional on participation. This approach is also shared by the literature on estimation of preferences with betting data.

the general identification strategy when the trader population exhibits heterogeneity in both risk preferences and beliefs.

To characterize equation (7) under our model of risk neutrality and heterogeneous beliefs, consider an generic race with  $n$  horses. Let us return to Lemma 1, which allows us to represent an agent  $t$ 's preferences in terms of the relative deviations from correct beliefs  $\mathbf{r}_t = (r_{1t}, \dots, r_{nt}) > 0$  for any agent  $t \in T$ . The model implies that agent  $t$  invests his/her endowment in security  $i = 1, \dots, n$  if

$$\frac{ER_i}{ER_j} \geq \frac{r_{it}}{r_{jt}} \quad \forall j \neq i.$$

Assuming a continuous distribution of  $\mathbf{r}_t = (r_{1t}, \dots, r_{nt}) > 0$  in the population (so that indifference between two securities has measure zero), risk neutrality gives rise to a discrete choice model for asset demand consistent with equation (7). The fact that the discrete choice behaviour is without loss of generality in our belief heterogeneity model is a desirable feature of our approach and not shared by the existing preference-based literature, which must impose discrete choice behaviour as an additional assumption.

Because choices are invariant to a monotonic transformation of preferences, we can express agent  $t$ 's utility for betting on asset  $i$  in the market as

$$U_t(p_i, R_i) = u_{it} = \delta_i + v_{it}, \quad (8)$$

where  $\delta_i = \log ER_i - \log ER_1$  and  $v_{it} = \log r_{1t} - \log r_{it}$ . Notice the utility of horse 1 has been normalized so that  $u_{1t} = 0$ . Given this utility specification, it is straightforward to check that  $U_t(p_i, R_i) > U_t(p_j, R_j)$  is equivalent to  $\frac{ER_i}{ER_j} \geq \frac{r_{it}}{r_{jt}}$ . Importantly,  $v_t = (v_{1t}, \dots, v_{nt})$  fully characterizes an agent's beliefs. To see why, observe that  $\log p_i - \log p_1 + v_{it} = \log(\pi_{it}) - \log(\pi_{1t})$  for each  $i = 2, \dots, n$ , where recall  $\pi_{it}$  is agent  $t$ 's subjective belief that horse  $i$  wins. Combined with the constraint that  $\sum_{i=1}^n \pi_{it} = 1$ , we can uniquely recover the vector of subjective beliefs  $\pi_t = (\pi_{1t}, \dots, \pi_{nt})$  from  $v_t$  (we have  $n$  equations in  $n$  unknowns). Thus, we can represent an agent's beliefs in terms of the vector of additive errors  $v_t$ .

Let  $P(\cdot; \theta)$  be the distribution of  $v_t$  that captures the belief distribution in the population.  $P$  is indexed by an unknown vector of parameters  $\theta \in \Theta$  that is the object of estimation. Assuming that  $P$  is continuous and that the distribution of endowments is independent of  $P$ , by Lemma 1 we can express equation (7) as

$$s_i = \int_{v_t \in \mathbb{R}^n} \mathbf{1}[\delta_i + v_{it} > \delta_j + v_{jt} \forall j \neq i] dP(v_{1t}, \dots, v_{nt}; \theta) \quad i = 1, \dots, n. \quad (9)$$

It turns out that this system of equations (9) is isomorphic to a *horizontally* differentiated product market in which each agent chooses among  $n$  products, with  $\delta_i$  representing the “mean utility” of product  $i$ —the mean utility of product 1 is normalized to be zero—and  $v_t = (v_{1t}, \dots, v_{nt})$  is a vector of random utility terms that is heterogeneous in the population—the random utility term  $v_{1t}$  of product 1 also being normalized to be zero. Inverting systems of equations like (9) is central to the demand estimation framework put forth by [Berry et al. \(1995\)](#) (henceforth referred to as BLP), *i.e.* solving for mean utilities given the underlying distribution of preferences over a set of differentiated products using the observed prices and market shares over these products. This connection provides us an immediately useful result: so long as  $P$  is continuous, then [Berry \(1994\)](#) shows that there exists a vector of mean utilities  $(\delta_1, \dots, \delta_n)$  that solves equation (9). Furthermore, the more recent [Berry et al. \(2012\)](#) shows that the solution is unique.

We can now exploit these results to carry out the empirical strategy of JS in a heterogeneous agents context. Constructing the likelihood for given choice of parameters  $\theta$  can be described by the following steps.

- (i) For any market  $k$ , we numerically find the unique underlying vector of mean utilities  $\delta^k = (\delta_1^k, \dots, \delta_{n^k}^k) \in \mathbb{R}^n$  that solves the system of equations (9).
- (ii) Given  $\delta^k = (\delta_1^k, \dots, \delta_{n^k}^k)$ , we can recover the underlying probability distribution over states of nature  $\mathbf{p}^k$  by using the following facts

- (a)  $\delta_i^k = \log ER_i^k - \log ER_1^k$ ,
- (b) The expected gross return is  $ER_i^k = (1 - \tau) \frac{p_i^k}{s_i^k}$ ,
- (c) Probabilities  $(p_1^k, \dots, p_{n^k}^k)$  over the  $n^k$  possible states of nature sum up to 1.

## 6. RESULTS

The only aspect of the estimation that remains to be discussed is how to parameterize  $P(v_i; \theta)$ . To discipline the parametric model we employ, we impose that belief heterogeneity is *idiosyncratic*. That is, we restrict  $P$  to satisfy  $E[v_{it}] = 0$  for all  $i = 1, \dots, n$ , and

$$E[v_{it} | v_{-i,t}] = 0, \quad (10)$$

for any  $v_{-i,t} \in \mathbb{R}^{n-1}$ . We interpret equation (10) as an important restriction for the random utility terms  $v_{it}$  to represent belief heterogeneity rather than preference heterogeneity. Under preference heterogeneity, if agent  $t$  has a large preference shock for one horse  $j$  (*i.e.* a high  $v_{jt}$ ) it could mean that he/she also has a high  $v_{it}$  for another horse  $i$  with similar characteristics as  $j$  in the race. Thus, even if  $E(v_{it}) = 0$ , preference heterogeneity would not necessarily satisfy (10). Accordingly, condition (10) prevents agents from exhibiting systematic tastes for horses, which could possibly create an artificial demand for longshots.

We introduce now a rich yet tractable way to model belief heterogeneity that respects idiosyncratic heterogeneity. In particular, we let the random utility terms  $v_{it}$  to be distributed according to a *variance mixture* of logistic errors. That is,

$$P(v_{2t}, \dots, v_{nt}; \theta) = \int \prod_{i=2}^n F(v_{it} | \sigma) dG(\sigma),$$

where  $G(\sigma)$  is the mixing distribution. This mixture, which appears new to the discrete choice literature, retains the key properties of idiosyncratic heterogeneity equation (10).<sup>10</sup>

We show in the Appendix (Section D) that the distribution  $G$  can be non-parametrically identified from the data. Here we discuss how  $G$  is estimated. Solving for the inner integral in equation (9) analytically (see *e.g.* Train, 2003) yields the well-known mixed logit demand

$$s_i = \int_{\sigma \in \mathbb{R}_+} \frac{\exp\left(\frac{1}{\sigma} \delta_i\right)}{\sum_{j=1}^n \exp\left(\frac{1}{\sigma} \delta_j\right)} dG(\sigma) \quad i = 1, \dots, n, \quad (11)$$

where the mixing takes place over the distribution of the variances  $\sigma$ . Although this system equation (11) bears a resemblance to the mixed logit demand that was originally estimated by BLP, there is a critical difference. In contrast to the usual mean mixture—which violates equation (10),

10. Conditional on  $\sigma$ ,  $F(v_{2t}, \dots, v_{nt} | \sigma)$  represents i.i.d. logit errors with common variance  $\sigma$  and thus satisfies idiosyncratic heterogeneity. Then, by integrating out over  $\sigma$ , idiosyncratic heterogeneity is retained.

TABLE 1  
Parameter estimates

Parameter	One-type estimate (std. error)	Two-types estimate (std. error)	Three-types estimate (std. error)	Two-types CARA estimate (std. error)
$\sigma_1$	0.067 (0.0035)	0.028 (0.0033)	0.014 (0.0058)	0.034 (0.0074)
$\sigma_2$	—	0.503 (0.0588)	0.690 (0.1757)	0.514 (0.0862)
$\sigma_3$	—	—	0.075 (0.0306)	—
$P_1$	1	0.716	0.461 (0.1237)	0.706
$P_2$	—	0.284 (0.0184)	0.219 (0.0390)	0.294 (0.020)
$P_3$	—	—	0.320	—
CARA( $\gamma$ )	—	—	—	0.003 (0.0031)
Log-likelihood	-307,391.5	-307,291.8	-307,291.3	-307,288.0
LR test <sup>a</sup>	199.4	—	1.0	7.6
( <i>p</i> -value)	(< 0.0001)	—	(0.6065)	(0.0058)

<sup>a</sup>The LR test statistic compares the model against the two-type specification and is given by  $2\ln LL_a - 2\ln LL_0$ , where  $LL_0$  and  $LL_a$  represent the log-likelihood of the null and the alternative model, respectively (where the alternative model nests the null).

our heterogeneity is governed by a variance mixture, thereby introducing a random coefficient  $1/\sigma$  on the mean utility term  $\delta_i$ . This random coefficient does not affect the existence and uniqueness of a solution to the system, but it does affect our ability to compute it because the contraction mapping proposed in BLP is no longer valid (the mean utilities cannot interact with random coefficients in their set-up). Instead, we minimize the sum of squared errors between observed shares and predicted shares using a quasi-Newton procedure with zero mean utilities as initial values.

A one-component mixture, *i.e.*  $G(\sigma)$  with only one point in its support, corresponds to the standard logit. Adding components gives rise to a finite mixture. We view this as a natural way to capture heterogeneity in the population as the different components can be interpreted as trader “types”. A finite mixture with  $J$  components corresponds to a parameter vector  $\theta = (\sigma_1, \dots, \sigma_J; P_1, \dots, P_J)$ , where  $\sigma_j$  for  $j=1, \dots, J$  are the support points of the finite mixture and  $P_j$  is the probability mass of component  $j$ .<sup>11</sup> The parameter estimates for a one-, two- and three-component mixture are presented in Table 1.

The first thing to note is that the two-component specification provides a much better fit than the one-component specification. To see this fact, we illustrate in Figure 3 the pattern of (net) expected returns implied by the one-component and the two-component specifications, respectively, which are obtained by regressing the predicted  $ER_i^k$  on  $\log s_i^k$  for all securities  $i$  and markets  $k$  in the data. While the one-component model does a poor job fitting the expected returns of longshots, the two-component model fits the observed pattern remarkably well. This improvement is further confirmed by the likelihood ratio (LR) test, which strongly rejects the one-component in favour of the two-component model.

Table 1 also shows that adding a third component does not significantly improve the log likelihood. In addition, it roughly reproduces the two-component model by splitting the

11. Any distribution over a compact support can be approximated by a multinomial distribution. See *e.g.* Theorem 10.15 in Aliprantis and Border (2006).

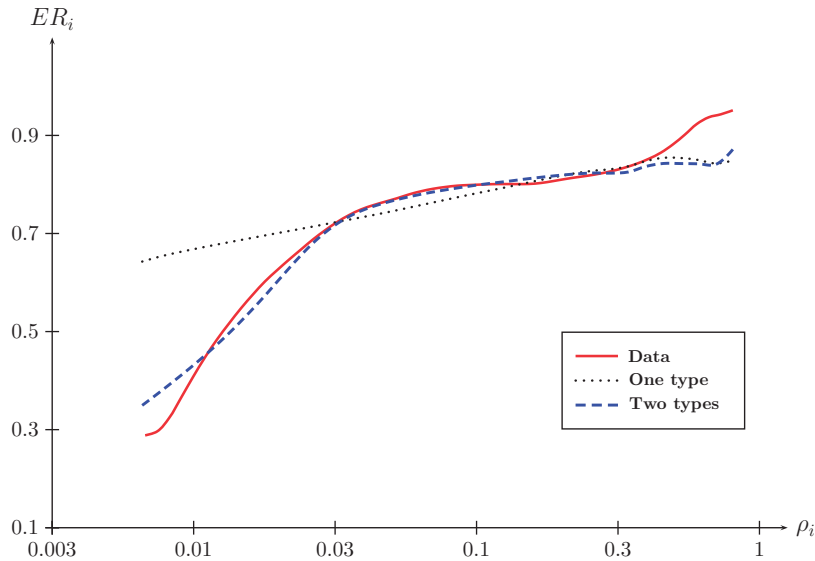


FIGURE 3  
Predicted and actual returns

predominant component ( $P_1=0.716$ ) into two low-variance types (1 and 3) with the second component still exhibiting substantial dispersion. Consequently, in what follows, we focus on the two-type model.

To check our assumption of risk neutrality, we also estimate a two-component model in which agents exhibit (homogeneous) constant absolute risk aversion (CARA) following the general approach laid out in the Appendix. The last column of Table 1 shows the model estimates. The estimated CARA coefficient (denoted by  $\gamma$ ) yields a tiny degree of risk aversion ( $\gamma=0.003$ ) albeit not significantly different from zero. In addition, the remaining parameters are virtually identical to the two-component risk-neutral version of the model, implying that the economic implications of these two models are identical. Nonetheless, allowing for CARA preferences yields a statistically significant likelihood gain.

### 6.1. Economic interpretation: informed and noise traders

The two-component mixture suggests a bimodality in the underlying distribution of agents in the market. In the context of financial markets, there is a large literature that theoretically introduces two types of agents: *arbitrageurs* who are relatively well informed, and traders exhibiting dispersed beliefs due to poor information or imperfect belief updating. This two-type classification is commonplace in the market microstructure literature (see *e.g.* [Glosten and Milgrom, 1985](#); [Kyle, 1985](#)), which separates traders into “noise” traders and insider speculators, in the “limits of arbitrage” literature ([De Long \*et al.\*, 1990](#); [Stein, 2009](#)), and in research related to mispricing in speculative markets ([Serrano-Padial, 2012](#); [Shin, 1991, 1992](#)). Although a pure noise trader in the literature is an agent that invests randomly, we use the term noise trader in the spirit of [De Long \*et al.\* \(1990\)](#), *i.e.* to refer to traders exhibiting belief dispersion. [Ottaviani and Sorensen \(2006\)](#) interpret these types in betting markets as *late* (arbitrageurs) and *early* traders (recreational gamblers).



If we view the components of the mixture as representing different “types” of agents, then type  $\sigma_1$  represents informed traders, which is also the modal type in the population (72%), while type  $\sigma_2$  is a noise trader, since it exhibits a much higher dispersion in subjective valuations. This is, to our knowledge, the first article empirically suggesting the existence of informed and noise traders using only aggregate data.<sup>12</sup>

To quantify the dispersion of type- $\sigma_2$  agents, consider the choice between investing in a security at random, yielding average returns in expectation, and an extreme longshot. For example, in our data set, the average (gross) return is 0.76, whereas it is  $\sim 0.35$  for securities in the bottom percentile of the price distribution. The probability that a type- $\sigma_2$  agent prefers a security with average returns to a security  $i$  yielding 0.35 is given by

$$Pr(0.76/0.35 > \bar{r}_i/r_{it}) = Pr\left(\bar{v}_i - v_{it} < \frac{\log 0.76 - \log 0.35}{0.503}\right) \approx 0.82,$$

where the last equality follows from the fact that  $\bar{v}_i - v_{it}$  is distributed standard logistic.

Furthermore, since type  $\sigma_2$  only represent 28% of the population, and type  $\sigma_1$  prefer the favourite virtually 100% of the time, the probability of the random investment being preferred to the longshot in the population as a whole is approximately  $0.82 \times 0.28 + 0.72 \approx 0.95$ . This is in stark contrast with a representative agent, who must be indifferent between the two bets to explain the data.

We further compare the representative agent and traders in our estimated model by looking at the distribution of indifference ratios  $\bar{r}/r_\rho$  in the two-component population. Figure 4 shows different quantiles of the distribution, which we compare alongside the implied indifference ratios of a representative agent. Agents in the interquartile range, *i.e.* the inner 50% of the agents in the population, exhibit very little dispersion from the canonical agent, *i.e.* the agent with indifference ratios = 1. That is, the modal behaviour in the heterogeneous population is closely captured by a risk-neutral agent with correct beliefs.<sup>13</sup> This is driven by the fact that the first component has a very low variance. Beliefs become more dispersed for agents outside the interquartile range, reflecting the fact that the variance of the second component is much higher ( $\sigma_2 = 0.503$  versus  $\sigma_1 = 0.028$ ). It is only at the far tails of our estimated distribution that we approach the extreme beliefs of the representative agent with respect to longshots.

Figure 4 also provides intuition for why the two-type specification performs better than the one-type. Consistent with the numerical example in Section 3, to generate the observed disparity in returns between longshots and the average bet, we need only a small fraction of traders investing in the longshot. This is because the empirical winning probabilities of longshots are very small. Accordingly, most of the agents belonging to the first-component exhibit beliefs that induce them to invest on the favourites, while agents in the second component spread their investment across securities, with a minority of them going for the extreme longshots. In contrast, the one-type specification faces a trade-off: either it exhibits low-belief dispersion (as it is the case) thus fitting well the returns on securities with moderate and high pay-out probabilities but does not generate enough demand for the longshots, or it exhibits higher dispersion, hence fitting the returns on favourites poorly.

12. There are several empirical studies that estimate the presence of noise traders using individual trading data from experimental markets (Cipriani and Guarino, 2005; Cowgill *et al.*, 2009; Forsythe *et al.*, 1992), or transaction data from financial markets (Easley *et al.*, 1997).

13. The interquartile range is a standard measure of dispersion and modal behaviour. Manski (2004) also uses it to document dispersion in beliefs.

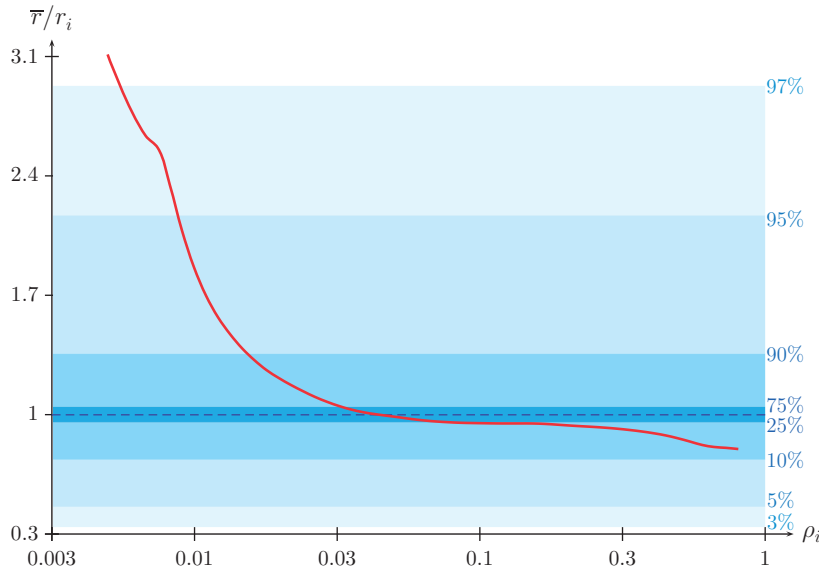


FIGURE 4  
Individual versus market behaviour

## 7. PREFERENCES VERSUS BELIEF HETEROGENEITY

We now contrast our belief heterogeneity (henceforth BH) approach with preference-based explanations of the FLB in two different ways. First, we formally compare the empirical fit of our estimated model to the standard preference explanation in the literature, probability overweighting, which JS and [Snowberg and Wolfers \(2010\)](#) have shown to explain the data better than risk loving. Secondly, we exploit the presence of exogenous variation in the information structure across races to non-parametrically test the predictions of our belief heterogeneity hypotheses against the predictions from a general class of preference-based theories.

We engage in formal model comparisons using the representative agent model of CPT preferences as was employed by JS and assess its performance, relative to our model, in terms of explaining the variation in the data. The CPT model we estimate is the preferred specification in JS: a three-parameter model consisting of a CARA utility function with risk coefficient  $\gamma$  for both monetary gains and losses, and two separate probability weighting functions: one for gains and the other for losses—this specification is richer than the one-parameter rank-dependent utility of [Snowberg and Wolfers \(2010\)](#). As in JS, we assume that the representative agent has correct beliefs about  $\mathbf{p}$ . The value of investing  $w$  on security  $i$  given odds  $R_i$  is given by

$$U(p_i, R_i) = G(p_i)u(wR_i, \gamma) + H(1 - p_i)u(-w, \gamma),$$

where  $u(x, \gamma)$  is a CARA utility,  $G(p) = p^\alpha$  is the weighting function for gains and  $H(p) = p^\beta$  is the weighting function for losses.<sup>14</sup> We estimate the model by maximum likelihood using indifference conditions (6)—see JS for details.

Our parameter estimates along with the CPT estimates from JS are presented in the first two columns of Table 2. As can be seen, both estimates are quite close, but because of the much

14. Observe that a single initial wealth level does not have to be assumed here because the CPT model considers only the utility of gains and losses relative to existing wealth.

TABLE 2  
CPT estimates

Parameter	Full sample estimate (std. error)	JS estimate (std. error)	Non-maiden estimate (std. error)	Maiden estimate (std. error)
CARA ( $\gamma$ )	-0.032 (0.0006)	-0.072 (0.021)	-0.020 (0.0029)	-0.032 (0.0041)
Gains ( $\alpha$ )	1.22 (0.0307)	1.162 (0.143)	1.12 (0.0307)	1.20 (0.0420)
Losses ( $\beta$ )	0.28 (0.0006)	0.318 (0.272)	<b>0.55</b> (0.0898)	<b>0.25</b> (0.0769)
Observations	176,466	4,037	87,394	29,003
Log-likelihood	-307,301.5	-7,365.3	-156,101.0	-50,508.8
Vuong Test ( $p$ -value)	34.92 ( $< 0.0001$ )			

Bold numbers represent significant difference between non-maiden and maiden estimates at the 1% level.

larger size of our sample, our estimates have a much higher precision. We find that preferences exhibit slight risk loving ( $\gamma < 0$ ), a slightly convex weighting of gains ( $\alpha > 1$ ) and a highly concave weighting of losses ( $\beta \ll 1$ ). Our estimates reinforce the major empirical finding from JS: under CPT, the key primitive driving the overpricing of longshots is the overweighting of loss probabilities (particularly for small probabilities).

### 7.1. Testing the models

Observe that the likelihood of our model exceeds that of the CPT model. This suggests that our beliefs approach exceeds the explanatory power of the existing CPT explanation. We can formalize this comparison in light of our identification strategy. Let  $\phi(\mathbf{R})$  denote the *true* mapping between prices and fundamentals in the data, *i.e.*  $\mathbf{p} = \phi(\mathbf{R})$ , and denote  $\phi^m(\mathbf{R}; \theta)$  the mapping between prices and fundamentals predicted by model  $m \in \{BH, CPT\}$  for  $\theta \in \Theta^m$ . If model  $m$  is properly specified, then

$$\phi(\mathbf{R}) = \phi^m(\mathbf{R}; \theta^m) \quad (12)$$

for some “true value” of parameters  $\theta^M \in \Theta^M$ . A natural question is whether equation (12) is actually testable, *i.e.* whether we can test that model  $m$  is consistent with the underlying data-generating process. The likelihood framework naturally allows for such a specification test, which was first proposed by White (1982). This specification test is derived from the information matrix equality, which is a fundamental theorem in likelihood theory (see *e.g.* Cameron and Trivedi, 2005). In our context, the information equality states that, if equation (12) holds, the following must be satisfied:

$$E \left[ \frac{\partial^2}{\partial \theta \partial \theta'} \log \phi_{i_{win}}^m(\mathbf{R}; \theta^m) \right] = -E \left[ \frac{\partial}{\partial \theta} \log \phi_{i_{win}}^m(\mathbf{R}; \theta^m) \frac{\partial}{\partial \theta} \log \phi_{i_{win}}^m(\mathbf{R}; \theta^m)' \right], \quad (13)$$

where the expectation is taken with respect to the true data-generating process over winners and odds ( $i_{win}, \mathbf{R}$ ) in the data. White’s test uses the sample analogues of each side of the equality to construct a quadratic form that should be sufficiently close to zero if equation (12) is true. We use the form of the test given by White (1987). Specifically, for each race  $k$  in the data let  $l_k(\hat{\theta}^m) = \log \left( \phi_{i_{win}}^m(\mathbf{R}^k; \hat{\theta}^m) \right)$ , *i.e.* the log likelihood value associated with race  $k$  at the estimated

parameters  $\hat{\theta}^m$ . The test statistic is constructed as

$$T^m = \left\{ K^{-1/2} \sum_{k=1}^K \hat{q}_k \right\}' \left\{ \frac{1}{K} \sum_{k=1}^K \hat{q}_k \hat{q}_k' \right\}^{-1} \left\{ K^{-1/2} \sum_{k=1}^K \hat{q}_k \right\},$$

where

$$\hat{q}_k = \text{vech} \left\{ \frac{\partial^2 l_k}{\partial \theta \partial \theta'} (\hat{\theta}^m) + \frac{\partial l_k}{\partial \theta} \frac{\partial l_k}{\partial \theta'} (\hat{\theta}^m) \right\}.^{15}$$

If equation (12) holds,  $T^m$  is asymptotically distributed  $\chi^2$  with  $J(J+1)/2$  degrees of freedom, where  $J$  is the length of the parameter vector  $\theta^m \in \Theta^m$ . Hence, large values of  $T^m$  indicate a significant difference between the model's predictions and the data.

The results of the test are rather stark: it fails to reject equation (12) for the BH model at standard significance levels ( $T = 8.54$ , whereas the 10% critical value is 10.64). Hence, we cannot reject the null hypothesis that the BH model equals the true data-generating process. In contrast,  $T^{CPT} = 1139.08$ , which clearly rejects the CPT model. Hence, although the log-likelihood values of the two models are somewhat close, as far as the information equality test is concerned there is a big difference in terms of their ability to accurately describe the true data-generating process. We do not interpret this test as literally telling us that our model is entirely properly specified. However, we do interpret it as saying that it captures the subtle variation in the data well enough such that it cannot be distinguished from the true data-generating process with nearly 200,000 races and only 3 parameters! We reach similar conclusions if instead we implement the non-nested model selection tests proposed by [Vuong \(1989\)](#) between the BH and CPT models (see the bottom row of Table 2).

## 7.2. Information differences and the FLB

We have shown that our estimated two-type model of belief heterogeneity better explains the FLB as compared to a representative agent CPT model. However, these model comparison tests have a few drawbacks. First, we have only offered one parametric alternative (the CPT model estimated by JS), which is far from exhausting the space of possible preference alternatives, including possibly heterogeneous preference models such as the one studied by [Chiappori et al. \(2012\)](#). Moreover, although we derived our random utility (equation 11) on the basis of belief heterogeneity, and our estimates have a natural interpretation in terms of informed and noise traders, such heterogeneity could also have a preference interpretation.

In light of this, how can we distinguish between our belief-based explanation and preference-based models? The key difference is the following: in a standard preference-based approach, an individual has preferences for binary gambles  $(p, R)$  and thus each individual's preferred asset in a race depends deterministically on the menu of gambles  $\{(p_i, R_i)\}$  in the race. In our belief-based model, this is no longer true: the true underlying menu of gambles is insufficient to characterize an individual's asset demand because an individual will have idiosyncratic deviations in beliefs. These belief shocks will vary for the same individual across both horses and races. Thus, for the same menu of gambles  $\{(p_i, R_i)\}$  in two different races, an agent will have different preference ranking over horses  $i = 1, \dots, n$  in the two races depending on the belief shocks he/she receives,

15.  $\text{vech}(A)$  denotes the half-vectorization of symmetric matrix  $A$ .

whereas in preference-based models the individual will have exactly the same preferences ranking over horses as long as the underlying menu of gambles is invariant across races.

This distinction has a very important testable implication: beliefs change with information, whereas preferences for gambles do not. Thus, the equilibrium prices predicted by our belief-based model will be different for two races with the same fundamentals but different public information, whereas the predictions from a preference-based model will be the same. This immediately suggests a non-parametric approach to testing between our beliefs model and preference alternatives—an ideal experiment would be to use exogenous variation in the amount of information about participating horses and empirically study whether this impacts pricing in the market. Unfortunately, our data do not contain any horse-specific information. However, we exploit the fact that horse races on the same day at the same track come in two different forms: *maiden* and *non-maiden*. Horses in a maiden race are those that have yet to win a single race, and thus by definition any new horses are entered into maiden races, whereas horses must have a racing history to participate in a non-maiden race. Thus, on average, there is much less handicapping information about horses in maiden races as compared to non-maiden (Camerer, 1998; Mitchell, 1989). Accordingly, for a given vector of winning probabilities  $\mathbf{p}$  and the same trader population, the lower uncertainty surrounding non-maiden races should lead to less belief dispersion and, in turn, to a less pronounced FLB, since traders can condition their beliefs on richer information—a fact we discuss in more detail below. In contrast, a purely preference-based model would yield identical predictions across race types.

To compare these predictions, we control for other differences that might exist between maiden and non-maiden races by restricting our samples to *claiming* races, in which horses can be purchased before the races and exhibit horses with similar price tags (thus trying to ensure a level playing field). This is the most frequent type of races in the U.S. (over 54% of all races). Furthermore, maiden and non-maiden claiming races take place typically on the same day at the same track,<sup>16</sup> and have similar track takes.<sup>17</sup> Thus, arguably the trader population should be similar across maiden and non-maiden claiming races.

Figure 5 shows the relationship between prices and returns for each race type. As can be seen, there are considerable price differences. Specifically, the magnitude of the FLB is much more pronounced in maiden compared to non-maiden races: while expected returns in maiden races go from \$0.05 for extreme longshots to \$0.96 for heavy favourites, returns in maiden races are much more compressed, ranging from \$0.45 to \$0.89, with the reduction in mispricing being especially prominent for longshots.

This result hints at a basic inconsistency of a preference-based view of the FLB: it can only explain pricing differences across types of races through variation in preferences. We illustrate the magnitude of this inconsistency by estimating the above CPT model on the separate subsamples. The estimates are given in the last two columns of Table 2, which show significant differences across samples, in particular on the probability weighting of losses (significant at the 1% level): while both the full sample and maiden parameters are very similar,  $\beta$  is more than twice as high in non-maiden races, resulting in a much less concave weighting function (see Figure 6). These large swings in preferences across maiden and non-maiden races are difficult to rationalize and, given the stark differences in prices across races, such swings will also be present in any alternative preference-based model.<sup>18</sup>

16. In our data set, 99.5% of maiden races were run in a day where non-maiden races were also run.

17. The average track take for maiden claiming races is 0.190 with std.dev. 0.04. For non-maiden claiming races the average take is 0.192 and the std.dev. is 0.04.

18. Following Chiappori *et al.* (2009), we also estimated a model with heterogeneous risk attitudes and found that while most agents are risk lovers in maiden races, they become risk averse in non-maiden races.

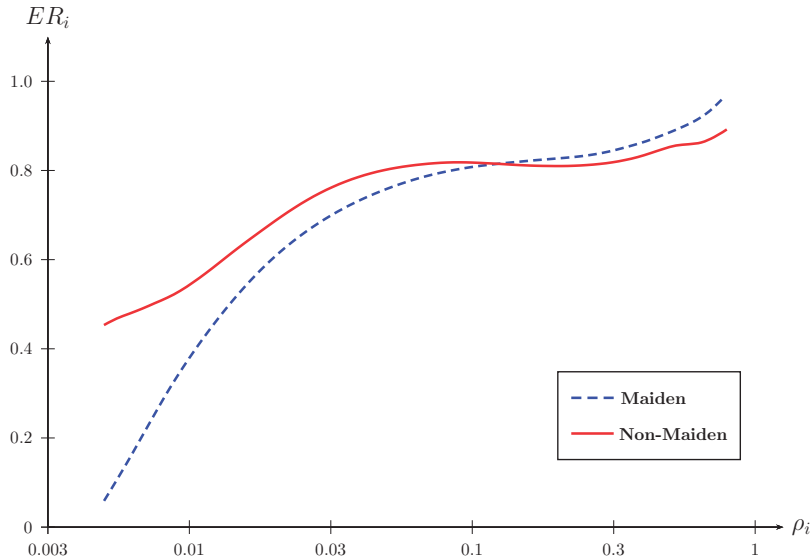


FIGURE 5  
The FLB in maiden and non-maiden races

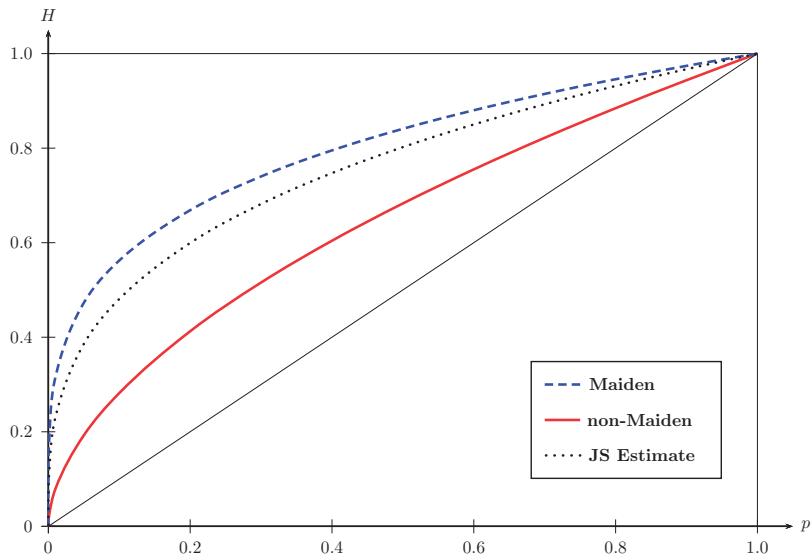


FIGURE 6  
CPT weighting function for losses

**7.2.1. Self-selection into races.** We now show that this preference instability across race types is not easily explained by the self-selection of agents into races based on heterogeneous risk preferences. The hypothesis is that, in addition to making an optimal choice of which horse to bet *within* a race, agents with different risk preferences are also selecting *across* races in a way that accounts for the pricing differences between maiden and non-maiden races.

This self-selection hypothesis has a clear testable implication. If agents are self-selecting based upon *well-behaved* risk preferences, then we should not observe any significant arbitrage opportunities across race types that would allow a bettor to receive higher odds at higher win probabilities. This is because all standard risk preferences, including expected utility and CPT, satisfy first-order stochastic dominance, meaning that agents prefer bets that exhibit both lower prices (higher odds) and lower risk (higher winning probability). Thus, if such opportunities exist, bettors should exploit them by switching bets across race types until prices adjust to the point where this dominance is eliminated. We will call the existence of such an arbitrage a “profitable switch.”

The graph in Figure 5 already suggests a large presence of profitable switches in the data: longshots in non-maiden races pay higher average returns than similar longshots in non-maiden races. This pricing asymmetry can be exploited to construct a betting strategy that switches bets from maiden to non-maiden races and earns both higher odds at higher win probabilities. To construct such a strategy, we use our two-type specification to estimate the underlying winning probabilities  $\mathbf{p}^k$  for each race  $k$  in the data—recall that we fail to reject it as the true data-generating process, so true probabilities should be very close to the ones predicted by the model. We then check for every horse in a maiden race whether there is a horse in a “nearby” non-maiden race that dominates or is dominated by the maiden horse (where “nearby” means a race that takes on the same day at the same track, and “dominance” means an asset pays higher odds with a higher winning probability than another). Thus, whenever a maiden horse is dominated by a non-maiden one in a nearby race, our model suggests a profitable switch from maiden to non-maiden that takes advantage of the fact that similar longshots are more favourably priced in non-maiden races.

We find that 91.6% of maiden races have at least one violation of dominance, and 81.5% include a profitable switch to a non-maiden race according to our model, that is, a horse  $i$  with price  $\rho_j^{k'} < \rho_i^k$  and probability  $p_j \geq p_i$  for some horse  $j$  in a non-maiden race  $k'$  run in the same day at the same track. If we restrict attention to adjacent maiden and non-maiden races, *i.e.* maiden races immediately preceded or followed by a non-maiden race in the same track, the percentage of violations is still quite high: 63.5%, with 49.5% of them including a profitable switch to a non-maiden race. By comparison, the frequencies of dominance violations across maiden races are only 31.1% for same-day races and 22.3% for adjacent races. Consequently, our model detects pervasive violations of dominance across race types, compared to violations within type, which would appear strongly at odds with selection on preferences being the cause of the pricing differences between races.

The dominance violations flagged by our model translate into a significant arbitrage opportunity. The second row of Table 3 shows that using the flags to switch from maiden to non-maiden in the same day at the same track leads to a 7.3% increase in the probability of winning and a 5% increase for adjacent switches. Because every switch involves a cheaper bet in a non-maiden race, traders would pay lower prices at no cost to the overall probability of winning. This generates a large effect on the average returns to betting: as the first row in Table 3 illustrates, *ex post* returns increase by 29.2% (23.8% for adjacent races).

These estimates show that selection into races based upon risk preferences alone cannot explain the preference instability across maiden and non-maiden races since any such selection should exploit the substantial arbitrage opportunities present across race types, and thus reduce the observed pricing differences rather than create them. The only selection effect involving preference that could explain the price variation is one based upon *both* risk attitudes and non-pecuniary preferences. In particular, the price differences between maiden and non-maiden races requires individuals betting in maiden races to be less risk averse. This margin of selection may seem natural. However, to avoid the “no arbitrage” argument above, there must also be an

TABLE 3  
*Alternative betting strategy*

Variable	Same day		Adjacent	
	Estimate <sup>a</sup> (Robust std. error)	Relative Change	Estimate (Robust std. error)	Relative Change
$\Delta Ex\ post$ returns	0.0935** (0.0231)	29.19%	0.0715* (0.0429)	23.85%
$\Delta Winner^b$	0.0042** (0.0012)	7.33%	0.0030* (0.0023)	4.98%
$\Delta Log\ price$	-0.0468** ( $< 0.0001$ )		-0.0612** ( $< 0.0001$ )	
Observations	65,252		20,375	

<sup>a</sup>Std. errors are clustered by race. \*\* Means  $p < 0.01$ ; \* Means  $p < 0.1$ .

<sup>b</sup>One-sided test. Null Hypothesis:  $\Delta Winner < 0$ .

unexplained preference shock for betting in maiden races. One difficulty with this explanation is that, if there were non-pecuniary preference shocks, *e.g.* a thrill to betting on one's favourite horse, these are likely to be more pronounced for horses with a longer racing history rather than new horses.

**7.2.2 Can belief heterogeneity explain the difference?** We now show that heterogeneous beliefs provides a natural explanation for the pricing disparity between maiden and non-maiden races. Since there is less public information in maiden races, we should expect more belief dispersion in maiden races, particularly among noise traders, leading to a more pronounced FLB. Nonetheless, our beliefs model does still impose a restriction we can test: the change in the FLB across race types should be driven by a change in the belief dispersion of each type and *not* by a change in the proportion of trader types. If our model instead were to empirically require a higher prevalence of noise traders in the population to explain the change in the pattern of returns, it would exhibit the same parameter instability that the preference-based theories suffer from. We formulate this claim as two testable hypotheses.

**Hypothesis 1.** *The fraction of types is the same across maiden and non-maiden races.*

**Hypothesis 2.** *Belief dispersion of each type is higher in maiden races compared to non-maiden races.*

The basic insight behind Hypothesis 2 is that, if agents are Bayesian, their beliefs should get closer as the amount of (public) information about the underlying state of the world accumulates over time. For instance, in the context of horse racing, if agents observe the performance of horses over time, they would eventually agree on the (true) winning probabilities.<sup>19</sup> Accordingly, if beliefs get closer as more information becomes available, we should also expect a smaller FLB in non-maiden races. We show in the Appendix that this is the case in a model where the same

19. Belief convergence is formally treated in the literature on merging of opinions (Blackwell and Dubins, 1962; Gossner and Tomala, 2008; Kalai and Lehrer, 1994; Lehrer and Smorodinsky, 1996), and models of learning in competitive markets (Easley and Blume, 1982; Feldman, 1987; Townsend, 1978; Vives, 1993). Although the mentioned research on merging concerns the evolution of beliefs in the long run, Sandroni and Smorodinsky (1999) show that the speed of convergence can be quite fast.



TABLE 4  
*BH estimates for claim and non-claim races*

Parameter	Non-maiden estimate (std. error)	Maiden estimate (std. error)	Full sample estimate (std. error)
$\sigma_1$	<b>0.0001</b> ( $< 0.0001$ )	<b>0.0366</b> (0.0079)	0.028 (0.0033)
$\sigma_2$	<b>0.340</b> (0.0180)	<b>0.6599</b> (0.1403)	0.503 (0.0588)
$P_1$	0.700	0.727	0.716
$P_2$	0.300 (0.0104)	0.273 (0.0321)	0.284 (0.0184)
Observations	87,394	29,003	176,466
Log-likelihood	-156,090.0	-50,505.8	-307,291.8

Bold numbers represent significant differences between non-maiden and maiden estimates at the 1% level.

A–D security market is repeated over time. Specifically, we show that, for given  $\mathbf{p}$ , the FLB gets mitigated when agents can observe the history of *ex post* returns and agree on the *direction* of the belief updating.

Table 4 presents the model estimates for the two samples, which confirm Hypotheses 1 and 2. The fraction of type one is very stable around 70–73%—we fail to reject Hypotheses 1 at any standard significance level. In contrast, belief dispersion estimates are significantly smaller (at the 1% level) in non-maiden races:  $\sigma_1$  is virtually zero and  $\sigma_2$  is about half its value in maiden races. Thus, remarkably, our heterogeneous beliefs approach explains the change in prices across information structures in the theoretically predicted way.

Overall, these results give credence to the idea that belief heterogeneity is a major driver behind the FLB because, unlike preference theories, it is able to naturally explain large changes in pricing patterns across races with different information environments. It is important to stress that the disparity between maiden and non-maiden prices does not rule out alternative representative agent models, such as those in which preferences exhibit some form of ambiguity aversion, since one can regard bets on a maiden horse as more ambiguous than bets on a non-maiden horse. For instance, agent preferences could follow CPT for uncertainty in which the weighting function is replaced by a non-additive capacity (see [Tversky and Kahneman, 1992](#)). Nonetheless, any such model would still need to explain the differences across race categories by means of a change in beliefs, regardless of whether those beliefs are represented by additive probabilities or not. In addition, it would exhibit the same weaknesses as a model in which the representative agent has expected utility preferences but his/her beliefs are chosen *ad hoc* to rationalize the data.

## 8. CONCLUSION

We have shown in this article that allowing for belief heterogeneity in asset markets can reconcile model predictions with observed aggregate patterns without compromising the validity of standard behavioural assumptions such as weak risk aversion and expected utility. We have used A–D security markets as a general setting in which to illustrate these points. We have also showed that a belief-based model of trade in financial markets outperforms existing preference-based explanations of the FLB on several dimensions. We think heterogeneous beliefs have the potential to play an important role in empirical work in other institutional settings, such as insurance and credit markets, and we hope this article encourages future research in this area.

## APPENDIX

## A. OMITTED PROOFS

*Proof of Lemma 1.* Consider the maximization problem of agent  $t$  with endowment  $w_t > 0$  and beliefs  $(\pi_{1t}, \dots, \pi_{nt}) \gg 0$ . Given the state of the world  $(p_1, \dots, p_n)$  and market prices  $(\rho_1, \dots, \rho_n) \gg 0$ , the agent solves

$$\begin{aligned} \max_{(x_1, \dots, x_n) \in \mathbb{R}_+^n} \sum_{i=1}^n x_i \left( \frac{\pi_{it}}{\rho_i} - 1 \right) + w_t \\ \text{s.t.} \quad \sum_{i=1}^n x_i \leq w_t. \end{aligned}$$

The ratio  $\pi_{it}/\rho_i$  represents the *subjective* (expected) returns of security  $i$ , *i.e.* the return given agent's beliefs. If there is a security  $i$  such that

$$\frac{\pi_{it}/\rho_i}{\pi_{jt}/\rho_j} > 1, \quad \forall j \neq i, \quad (\text{A.1})$$

then the solution to the agent's problem implies investing all the endowment  $w_t$  in security  $i$ . This is because: (i) security  $i$  yields the strictly highest subjective returns among all securities; and (ii) security  $i$ 's subjective returns are strictly  $> 1$ . Property (ii) comes from the fact that, since  $\sum_i \pi_i = 1$  and  $\sum_i s_i = 1$ , we must have  $\max_h \frac{\pi_{ht}}{s_h} > 1$  whenever  $\frac{\pi_{ht}}{s_h} \neq \frac{\pi_{kt}}{s_k}$  for some  $h, k$ . This also means that, if the latter is true, the agent will invest all his/her endowment in the securities yielding  $\max_h \pi_{ht}/s_h$ , being indifferent about how much to invest on each of them. Finally, if  $\frac{\pi_{ht}}{s_h} = \frac{\pi_{kt}}{s_k}$  for all  $h, k$ , then subjective returns are all equal to 1, making the agent indifferent between investing any amount in  $[0, w_t]$ . To finish this case, notice that, by letting  $r_{it} = p_i/\pi_{it}$ , expression (A.1) becomes

$$\frac{ER_i}{ER_j} > \frac{r_{it}}{r_{jt}}, \quad \forall j \neq i. \quad \parallel$$

*Proof Theorem 1.* We consider first the case of a market with zero transaction costs  $\tau = 0$ . Let  $q_i(ER_1, \dots, ER_n; \mathbf{p}, \theta)$  represent the aggregate investment on asset  $i$  and denote  $\mathbf{r}_{it} = \left( \frac{r_{it}}{r_{1t}}, \dots, \frac{r_{it}}{r_{i-1t}}, \frac{r_{it}}{r_{i+1t}}, \dots, \frac{r_{it}}{r_{nt}} \right)$ . Given Lemma 1,  $q_i$  at fair prices satisfies, for all  $(\mathbf{p}, \theta)$ ,

$$\int_T w_t \mathbf{1}[\mathbf{r}_{it} < (1, \dots, 1)] dt \leq q_i(1, \dots, 1; \mathbf{p}, \theta) \leq \int_T w_t \mathbf{1}[\mathbf{r}_{it} \leq (1, \dots, 1)] dt.$$

By Assumption 1, the lower bound on  $q_i$  in this expression is bounded away from zero while the upper bound is bounded below  $\int_T w_t dt$ . In addition, aggregate investment is bounded by the aggregate endowment in the market:

$$\sum_j q_j(ER_1, \dots, ER_n; \mathbf{p}, \theta) \leq \int_T w_t dt.$$

This implies that, the market share of security  $i$  at fair prices, which is given by

$$s_i(1, \dots, 1; \mathbf{p}, \theta) = \frac{q_i(1, \dots, 1; \mathbf{p}, \theta)}{\sum_j q_j(1, \dots, 1; \mathbf{p}, \theta)},$$

must be bounded above zero and below one. Define

$$\underline{s} := \min_i \inf \{s_i(1, \dots, 1; \mathbf{p}, \theta) : (\mathbf{p}, \theta) \in (0, 1) \times \Theta\} > 0.$$

To prove part (i) of the theorem, let  $p_i > \bar{q}_i := 1 - \underline{s}$  and suppose the theorem does not hold, *i.e.*  $p_i \leq \rho_i = s_i$ . Then, since  $\sum_{j \neq i} s_j = 1 - s_i$  and  $\sum_{j \neq i} p_j = 1 - p_i$  there must be at least a security  $h$  with  $ER_h \geq 1$ . Pick the one with the highest expected returns. For this security we must have  $\frac{ER_h}{ER_j} \geq 1$  for all  $j$ , and thus it must be that  $s_h \geq \underline{s}$ . But then, since  $s_i \geq p_i > 1 - \underline{s}$  by assumption, we have that  $s_h \leq \sum_{j \neq i} s_j < \underline{s}$ , a contradiction. Obviously if security  $i$  is strictly underpriced, we must have that the remaining securities are overpriced on average:  $\sum_{j \neq i} \rho_j > \sum_{j \neq i} p_j$ .

Now consider the introduction of transaction costs. If, before introducing them, trader  $t$  was indifferent between investing on any security  $i$  or not investing,<sup>20</sup> we must have

$$\frac{ER_i}{ER_j} = \frac{r_{it}}{r_{jt}} \quad \forall j \neq i. \quad (\text{A.2})$$

20. Recall from the proof of Lemma 1 that for an agent to be indifferent between investing or not he/she must also be indifferent between investing among any two securities.

In this context, let  $ER_i = \frac{p_i}{s_i^*}$ , where  $s_i^*$  is the market share associated with zero transaction costs. After introducing positive transaction costs, for the agent to be indifferent between investing in security  $i$  and not investing in the market, expected returns need to satisfy

$$ER_i = (1 - \tau) \frac{p_i}{s_i'} = \frac{p_i}{s_i^*},$$

where  $s_i'$  is the new share for  $i$ , implying that  $s_i' < s_i^*$ . But this means that (i) the agent does no longer invest in the market at the old prices ( $\rho_i = s_i^*$  for all  $i$ ), and (ii) if he/she is indifferent between investing in  $i$  or not at the new prices, he/she has strict incentives not to invest in some of the other securities. This is because  $s_i' < s_i^*$  implies  $s_j' > s_j^*$  for some security  $j$ , given that market shares add up to one. Hence, the above indifference condition translates into

$$\frac{ER_i}{ER_j} \geq \frac{r_{it}}{r_{jt}} \quad \forall j \neq i, \quad (\text{A.3})$$

with strict inequality for some  $j$ .

Summing up, introducing transaction costs can potentially reduce the demand for any given asset. If this reduction is big enough the equilibrium market share for some security may not be bounded above zero, also implying that the other securities' shares may not be bounded below one, so that the FLB may not hold. Thus, to show that the FLB holds in equilibrium we need to show that this does not happen for low enough  $\tau$ .

The first thing to note is that, since  $L_i[z|\mathbf{p}, \theta] = Pr[\mathbf{r}_{it} \ll z|\mathbf{p}, \theta]$  and  $L_i[1|\mathbf{p}, \theta]$  is bounded away from zero, we can always find  $z = (z_1, \dots, z_n) \ll (1, \dots, 1)$  close enough to  $(1, \dots, 1)$  such that  $L_i[z|\mathbf{p}, \theta]$  is also bounded above from zero for all  $i$ .

Secondly, notice that the smaller the track take  $\tau$ , the closer the market share  $s_i'$  gets to  $s_i^*$  with  $s_i' \rightarrow s_i^*$  as  $\tau \rightarrow 0$  for all  $i$ , and thus the closer the indifference condition (A.3) gets to condition (A.2). Hence, if we fix  $ER_i/ER_j = 1$  for all  $i, j$ , given  $z \ll (1, \dots, 1)$  we can always find a low enough  $\bar{\tau} > 0$  such that, for all  $\tau < \bar{\tau}$ , the marginal traders indifferent between investing in  $i$  or not have indifference ratios satisfying  $z < \mathbf{r}_{it} < 1$ .

Therefore, combining these two facts we can show that market shares when  $ER_i/ER_j = 1$  for all  $i, j$  are bounded away from zero for all  $\tau < \bar{\tau}$  given some small  $\bar{\tau} > 0$  and apply the same reasoning as in the case of  $\tau = 0$  to show that the FLB must hold in equilibrium.  $\parallel$

### A.1. Weakening Assumption 1

As we mention in Section 3, when endowments and beliefs are independent, the FLB could obtain even when the mass of agents strictly preferring security  $i$  over the alternatives dwindles to zero as  $p_i \rightarrow 0$ . The next assumption and theorem formalize this intuition.

**Assumption 3.** *There exist  $\underline{p} > 0$  and  $\alpha > 1$  such that  $L_i[(1, \dots, 1)|\mathbf{p}, \theta] > \alpha p_i$  for all  $p_i < \underline{p}$  and all  $i = 1, \dots, n$ .*

**Theorem 3.** *If Assumption 3 holds and endowments are independent of  $P$ , there exists  $\bar{\tau} > 0$  such that for all  $\tau < \bar{\tau}$  a necessary consequence of equilibrium is that there exists  $\bar{q} < 1$  such that, for all  $i = 1, \dots, n$ , if  $p_i > \bar{q}$  then security  $i$  is underpriced while securities  $j \neq i$  are overpriced on average.*

*Proof* The same argument as the one used in the proof of Theorem 1 follows through by noticing that the orthogonality of endowments and beliefs implies that  $s_i$  is bounded below by  $L_i$ . To see why, focus on the case of  $\tau = 0$  and notice that

$$\begin{aligned} s_i &\geq \frac{1}{\int_T w_i dt} \int_T w_i \mathbf{1}\{r_{it}/r_{jt} < ER_i/ER_j \forall j \neq i\} dt \\ &= \frac{1}{\int_T w_i dt} \left( \int_T w_i dt \right) L_i[(\dots, ER_i/ER_j, \dots)|\mathbf{p}, \theta], \end{aligned}$$

where the last equality follows from the independence of endowments and indifference ratios. But then, the above assumption guarantees that  $s_i \geq L_i[(1, \dots, 1)|\mathbf{p}, \theta] > p_i$  at fair prices for all  $\mathbf{p}$  with  $p_i < \underline{p}$ . Given this, the same argument by contradiction used in the proof of Theorem 1 immediately applies.  $\parallel$

## B. PUBLIC INFORMATION AND THE FLB

In this section, we illustrate how the FLB is mitigated by the arrival of information, which is consistent with the above findings regarding maiden and non-maiden races.

Consider the following scenario. There is going to be a sequence of two A–D security markets involving the same securities—the same result readily generalizes to having more than two markets. True pay-out probabilities  $\mathbf{p}$  are the

same across markets. We assume that the first time the market is run, agents have heterogeneous posterior beliefs. We also assume that the distribution of those beliefs is continuous with full support in  $\text{int} \Delta^{n-1}$ ,  $\tau = 0$  and that, in the absence of public information, agents' inferences and equilibrium prices in both markets would be the same. Markets are identical, *except* in the amount of public information available: before trading in the second market, all agents observe the realized outcome in the first market, *i.e.* which security paid positive returns. Let  $\mathcal{H}_i$  denote the event that security  $i$  pays out in the first market. Given agent  $t$ 's beliefs, let  $\mathcal{L}_{ij}^t(\mathcal{H}_i)$  denote the likelihood ratio associated with outcome  $\mathcal{H}_i$ . That is,  $\mathcal{L}_{ij}^t(\mathcal{H}_i) = \frac{Pr^t(\mathcal{H}_i|i)}{Pr^t(\mathcal{H}_i|j)}$ , where  $Pr^t(\mathcal{H}_i|j)$  is agent  $t$ 's probability assessment of observing  $\mathcal{H}_i$  conditional on security  $j$  paying out in the second market. According to Bayes' rule, the ratio of agent  $t$ 's subjective probabilities conditional on observing  $\mathcal{H}_i$  satisfies

$$\frac{\pi'_i}{\pi'_j} = \mathcal{L}_{ij}^t(\mathcal{H}_i) \frac{\pi_{it}}{\pi_{jt}}, \quad i, j = 1, \dots, n, \quad (\text{B.4})$$

where  $\pi$  and  $\pi'$  represent beliefs before and after observing  $\mathcal{H}_i$ , respectively. Our next result shows that, whenever agents agree on the “direction” of the updating, the release of public information about past outcomes decreases the underpricing of favourites and the average overpricing of longshots, that is, the FLB is mitigated. First, we provide the formal notions of agreement and unbiased interpretation of information we use in the result.

**Definition 1.** Agents ordinally agree on the interpretation of  $\mathcal{H}_i$  if  $\mathcal{L}_{ij}^t(\mathcal{H}_i) < (>) 1$  for some  $t$  implies  $\mathcal{L}_{ij}^{t'}(\mathcal{H}_i) < (>) 1$  for all  $t'$  and all  $i, j$ . Agents ordinally agree if they agree on the interpretation of  $\mathcal{H}_i$  for all  $i = 1, \dots, n$ .

Notice that ordinal agreement is weaker than requiring beliefs to be concordant (Milgrom and Stokey, 1982; Ottaviani and Sorensen, 2012), which would imply ‘cardinal’ agreement, *i.e.*  $\mathcal{L}_{ij}^t = \mathcal{L}_{ij}^{t'}$  for all  $i, j, t$  and  $t'$ .

**Definition 2.** Agent  $t$  is unbiased if  $\mathcal{L}_{ij}^t(\mathcal{H}_i) > 1$  for all  $j \neq i$  and all  $i = 1, \dots, n$ .

The notion of unbiased beliefs implies that after observing  $i$  pay out a trader revises upwards his/her beliefs about  $i$  paying out in the second market relative to all other securities. This leads to both less underpricing of heavy favourites and less overpricing on average of longshots. That is, public information mitigates the FLB, as it is the case in Figure 5.

**Proposition 1.** If agents are unbiased and ordinally agree then there exists  $\bar{p}$  such that for all  $p_i > \bar{p}$  the expected price of security  $i$  in the second market is higher than in the first market, and the prices of securities  $j \neq i$  are lower on average.

This result implies that when Theorem 1 holds, new information mitigates the FLB: heavy favourites are less underpriced and longshots exhibit less overpricing on average.

*Proof* Fix the state of the world  $\mathbf{p} = (p_1, p_2, \dots, p_n)$ . For simplicity, we assume that, when indifferent between investing or not in the market, all agents decide to invest—the proof logic would be the same as long as the arrival of public information does not alter agents' decision to participate when indifferent. Given equation (B.4), market shares (*i.e.* prices) in the second market after observing  $\mathcal{H}_i$  are given by

$$\begin{aligned} s_i &= \frac{1}{\int_T w_t dt} \int_T w_t \mathbf{1} \left[ \left( \frac{\pi'_{it}}{\pi'_{1t}}, \dots, \frac{\pi'_{it}}{\pi'_{nt}} \right) < \left( \frac{s_i}{s_1}, \dots, \frac{s_i}{s_n} \right) \right] dt \\ &= \frac{1}{\int_T w_t dt} \int_T w_t \mathbf{1} \left[ \left( \frac{\pi_{it}}{\pi_{1t}}, \dots, \frac{\pi_{it}}{\pi_{nt}} \right) < \left( \mathcal{L}_{1i}^t(\mathcal{H}_i) \frac{s_i}{s_1}, \dots, \mathcal{L}_{ni}^t(\mathcal{H}_i) \frac{s_i}{s_n} \right) \right] dt. \end{aligned}$$

It is straightforward to check that, when the distribution of prior beliefs is continuous and has full support, the release of public information leads to a higher  $s_i$  when  $\mathcal{L}_{ij}^t(\mathcal{H}_i) > 1$  for all  $j$  and all  $t$ : at any given price, the mass of agents that would consider security  $i$  the optimal investment has gone up after observing the information.

Next, notice that we must have  $Pr(\mathcal{H}_i) \rightarrow 1$  as  $p_i \rightarrow 1$ . Thus, if agents ordinally agree and are unbiased then  $Pr(\mathcal{L}_{ij}^t(\mathcal{H}_i) > 1 \forall j \neq i) \rightarrow 1$  as  $p_i \rightarrow 1$ , implying that the probability that  $\rho_i (= s_i)$  is higher in the second market is close to one for  $p_i$  close to one. That is, for  $p_i$  sufficiently high the expected price of security  $i$  is higher in the second market than in the first market. This, in turn, implies that  $\sum_{j \neq i} \rho_j$  goes down in expectation.  $\parallel$

### C. GENERAL APPROACH TO ESTIMATING HETEROGENEITY

We describe here how to conduct Step 1 in the estimation of a heterogeneous population model laid out in Section 5, which involves solving the system of equations (7):

$$s_i^k = \frac{1}{\int_T w_t dt} \int_T w_t \mathbf{1} \left[ U_t(p_i^k, R_i^k) > U_t(p_j^k, R_j^k) \forall j \neq i \right] dt, \quad i = 1, \dots, n^k.$$

Accordingly, the goal is to characterize  $U_t$  so that we can estimate the heterogeneous population model by inverting the system of equations (7) to recover  $\mathbf{p}^k$ . To do so, we assume agents have expected utility preferences and exhibit heterogeneity in beliefs and in risk attitudes. Accordingly, agent  $t$ 's pay off from investing in security  $i$  is given by

$$U_t(p_i, R_i) = \pi_{it} u(w_t R_i, \gamma_t) + (1 - \pi_{it}) u(-w_t, \gamma_t), \quad (\text{C.5})$$

where  $\pi_{it}$  is agent  $t$ 's belief about security  $i$  and  $u$  is utility over wealth with risk attitudes governed by the one-dimensional parameter  $\gamma_t$ —e.g. the coefficient of absolute risk aversion in CARA utility. The cardinality of expected utility allows to normalize the utility from losing  $w_t$  to be zero, i.e.  $u(-w_t, \gamma_t) = 0$ . In addition, since agent preferences over securities in a given market are invariant to a monotonic transformation of utility, we can take the log of the RHS of equation (C.5) and subtract  $\log p_1$  to write  $U_t$  as

$$U_t(p_i, R_i) = \log u(w_t R_i, \gamma_t) + \xi_i + v_{it},$$

where  $\xi_i = \log p_i - \log p_1$  and  $v_{it} = -\log r_{it}$ . If we assume that risk parameters and beliefs are independently distributed and that agents have equal endowments (normalized to 1),<sup>21</sup> we can write equation (7) as

$$s_i^k = \int_{\gamma \in \mathbb{R}} \int_{v_t \in \mathbb{R}^n} \mathbf{1} \left[ \log u(R_i^k, \gamma_t) + \xi_i + v_{it} > \log u(R_j^k, \gamma_t) + \xi_j + v_{jt} \forall j \neq i \right] dP(v_t; \theta) dH(\gamma_t),$$

for all  $i = 1, \dots, n^k$ , where  $P(\cdot; \theta)$  is the (continuous) distribution of  $v_t = (v_{1t}, \dots, v_{n^k t})$  and  $H$  represents the distribution of risk parameter  $\gamma_t$ . Thus, as long as this system of market shares has a unique solution in terms of the vector  $(\xi_1, \dots, \xi_{n^k})$ , we can recover the underlying probabilities  $\mathbf{p}^k$  from observed odds  $\mathbf{R}^k$ .

One way to ensure the existence of a unique solution is to assume, as we do in our estimation, that  $v_t$  is a variance mixture of independent logit errors with mixing distribution  $G(\sigma)$ . Accordingly, market shares are given by

$$s_i^k = \int_{\gamma_t \in \mathbb{R}} \int_{\sigma \in \mathbb{R}_+} \frac{\exp\left(\frac{1}{\sigma} \log u(R_i^k, \gamma_t) + \frac{1}{\sigma} \xi_i\right)}{\sum_{j=1}^{n^k} \exp\left(\frac{1}{\sigma} \log u(R_j^k, \gamma_t) + \frac{1}{\sigma} \xi_j\right)} dG(\sigma) dH(\gamma_t), \quad i = 1, \dots, n^k. \quad (\text{C.6})$$

This system has a solution, given that it satisfies the sufficient conditions in Berry (1994), and it is unique as shown by Berry *et al.* (2012).

### D. NON-PARAMETRIC IDENTIFICATION OF BELIEF HETEROGENEITY

In this section, we formally prove identification of the distribution of beliefs, which is the key primitive we estimate in our model. The key source for identification stems from variation in the fundamentals  $\mathbf{p} = (p_1, \dots, p_n) \in \text{int} \Delta^{n-1}$  across different markets. To focus our attention on the logic of identification, we consider here a simplified setting with two horses and track take  $\tau = 0$ . Our discussion, however, easily generalizes to the  $n$ -horse context. Indeed, the additional variation made possible by  $n$ -horses rather than two-horses only aids identification rather than complicating it. Observe that an  $n$ -horse can always replicate a two-horse race by letting  $n-1$  horses have arbitrarily similar state probabilities but not vice versa, and thus the two horse problem is the essential setting to study.

In each race (i.e. market) we have two horses labelled  $i = 1, 2$  and can observe the market share  $s$  of horse 1 (the share of horse 2 being simply  $s_2 = 1 - s$ ) and expected returns  $(ER_1, ER_2)$  as a function of the market share. The belief heterogeneity model relates these market observables via the random utility

$$u_t = \log ER_1 - \log ER_2 - \epsilon_t,$$

where  $\epsilon_t$  corresponds to the belief disturbance associated with agent  $t$ . Agent  $t$  in the market bets on horse 1 if  $u_t > 0$ . Assuming a continuum of agents in the market and  $\epsilon_t$  distributed independently of  $(ER_1, ER_2)$  with distribution function  $H$ , we thus have that

$$s = H \left( \log \frac{p}{s} - \log \frac{1-p}{1-s} \right), \quad (\text{D.7})$$

since  $ER_i = \frac{p_i}{s_i}$ . We assume  $H$  satisfies the usual regularity conditions for a CDF, namely, it is continuous and strictly monotone over the support. Note that, for any  $s \in (0, 1)$  there exists a unique  $p$  that solves the above equation. In particular

21. Alternatively, we could assume  $w_t = a > 0$  for all  $t \in T$ . Generally, without information on endowments, the risk parameter is not separately identified from  $a$ .

letting  $p \rightarrow 1$  we have that  $\log \frac{p}{s} \rightarrow 0$  and  $\log \frac{1-p}{1-s} \rightarrow -\infty$ , and thus, the RHS of equation (D.7) approaches 1. Likewise as  $p \rightarrow 0$  we have that RHS of equation (D.7) approaches 0. Because the RHS of (D.7) is continuous in  $p$ , then by the intermediate value theorem for any  $s^c \in (0, 1)$  there exists a  $p^c \in (0, 1)$  that solves equation (D.7) and hence sustains  $s^c$  in equilibrium. Thus, the support of  $s$  under the model is the full unit interval  $(0, 1)$  and identification of  $H$  becomes straightforward: if we invert both sides of equation (D.7) we have

$$H^{-1}(s) = \log \frac{p}{s} - \log \frac{1-p}{1-s}. \quad (\text{D.8})$$

Because all the terms on the RHS are derived from the data, we thus identify the inverse of  $H$  for all  $s \in (0, 1)$  and hence  $H$  itself.

In the model we estimate, we restrict  $H$  further to be a variance mixture of logistic distributions. That is, for any  $t \in \mathbb{R}$  we model

$$H(t) = \int_{\sigma > 0} F(t|\sigma) dG(\sigma),$$

where  $F(\cdot|\sigma)$  is a standard logistic distribution with scale parameter  $\sigma$ . The object we estimate is  $G$ , *i.e.* the distribution of “types” in the population where each type corresponds to a different variance in beliefs, and  $G$  is assumed to have finite support. It is important to emphasize that the support points themselves are not known *ex ante*. Hence, if we let  $\mathcal{G}$  denote the set of all distribution functions with finite support, this is an infinite dimensional space of distributions. Using standard results from the identifiability of logistic mixture models (see Theorem 1 in Shi *et al.*, 2012), we can recover  $G$  from  $H$ .

*Acknowledgments.* This article was earlier circulated under the title “The Aggregate Effects of Idiosyncratic Heterogeneity: Evidence from an Arrow–Debreu Securities Market.” The authors are grateful to Enrique Sentana (Co-Editor) and three anonymous referees, whose comments greatly benefited the article, and to Ken Hendricks, Jean-Francois Houde, Marco Ottaviani, Antonio Penta, Dan Quint, Marzena Rostek, and Marek Weretka for very useful suggestions. They also thank Zhentong Lu for outstanding research assistance.

### Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

### REFERENCES

- ALI, M. (1977), “Probability and Utility Estimates for Racetrack Betting”, *Journal of Political Economy*, **85**, 803–815.
- ALIPRANTIS, C. D. and BORDER, K. C. (2006), *Infinite Dimensional Analysis: A Hitchhiker’s Guide* (Berlin, Germany: Springer).
- BERRY, S. T. (1994), “Estimating Discrete-Choice Models of Product Differentiation”, *The RAND Journal of Economics*, **25**, 242–262.
- BERRY, S. T., GANDHI, A. K. and HAILE, P. A. (2012), “Connected Substitutes and Invertibility of Demand” (Mimeo).
- BERRY, S. T., LEVINSOHN, J. and PAKES, A. (1995), “Automobile Prices in Market Equilibrium”, *Econometrica*, **63**, 841–890.
- BLACKWELL, D. and DUBINS, L. (1962), “Merging of Opinions with Increasing Information”, *Annals of Mathematical Statistics*, **33**, 882–886.
- BOMBARDINI, M. and TREBBI, F. (2012), “Risk Aversion and Expected Utility Theory: An Experiment with Large and Small Stakes”, *Journal of the European Economic Association*, **10**, 1348–1399.
- CAMERER, C. F. (1998), “Can Asset Markets Be Manipulated? A Field Experiment with Racetrack Betting”, *Journal of Political Economy*, **106**, 457–482.
- CAMERON, A. C. and TRIVEDI, P. K. (2005), *Microeconometrics: Methods and Applications* (Cambridge University Press).
- CHIAPPORI, P.-A., GANDHI, A., SALANIE, B. and SALANIE, F. (2009), “Identifying Preferences under Risk from Discrete Choices”, *American Economic Review: Papers & Proceedings*, **99**, 356–362.
- CHIAPPORI, P.-A., SALANIE, B., SALANIE, F. and GANDHI, A. (2012), “From Aggregate Betting Data to Individual Risk Preferences” (Mimeo).
- CIPRIANI, M. and GUARINO, A. (2005), “Noise Trading in a Laboratory Financial Market: A Maximum Likelihood Approach”, *Journal of the European Economic Association*, **3**, 315–321.
- COCHRANE, J. H. (2007), “Efficient Markets Today” (Chicago, Illinois: Conference on Chicago Economics).
- COWGILL, B., WOLFERS, J. and ZITZEWITZ, E. (2009) “Using Prediction Markets to Track Information Flows: Evidence from Google” (Mimeo).
- EASLEY, D. and BLUME, L. (1982) “Learning to be Rational”, *Journal of Economic Theory*, **26**, 340–351.
- EASLEY, D., KIEFER, N. M. and O’HARA, M. (1997) “One Day in the Life of a Very Common Stock”, *Review of Financial Studies*, **10**, 805–835.

- FELDMAN, M. (1987), "An Example of Convergence to Rational Expectations with Heterogeneous Beliefs", *International Economic Review*, **28**, 635–650.
- FORSYTHE, R., NELSON, F., NEUMANN, G. R. and WRIGHT, J. (1992), "Anatomy of an Experimental Political Stock Market", *American Economic Review*, **82**, 1142–1161.
- GLOSTEN, L. R. and MILGROM, P. R. (1985) "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders", *Journal of Financial Economics*, **14**, 71–100.
- GOLEC, J. and TAMARKIN, M. (1998), "Bettors Love Skewness, Not Risk, at the Horse Track", *Journal of Political Economy*, **106**, 205–225.
- GOSSNER, O. and TOMALA, T. (2008), "Entropy Bounds on Bayesian Learning", *Journal of Mathematical Economics*, **44**, 24–32.
- GRIFFITH, R. M. (1949), "Odds Adjustment by American Horse-Race Bettors", *American Journal of Psychology*, **62**, 290–294.
- JULLIEN, B. and SALANIÉ, B. (2000), "Estimating Preferences under Risk: The Case of Racetrack Bettors", *Journal of Political Economy*, **108**, 503–530.
- KALAI, E. and LEHRER, E. (1994), "Weak and Strong Merging of Opinions", *Journal of Mathematical Economics*, **23**, 73–86.
- KEIM, D. B. (2008), "Financial Market Anomalies", in Durlauf, S. N. and Blume, L. E. (eds), *The New Palgrave Dictionary of Economics* (Basingstoke: Palgrave Macmillan).
- KYLE, A. S. (1985), "Continuous Auctions and Insider Trading", *Econometrica*, **53**, 1315–1336.
- LEHRER, E. and SMORODINSKY, R. (1996) "Merging and Learning", *Lecture Notes-Monograph Series*, **30**, 147–168.
- LONG, J. B., SHLEIFER, A., SUMMERS, L. H. and WALDMANN, R. T. (1990) "Noise Trader in Financial Markets", *Journal of Political Economy*, **98**, 703–738.
- MANSKI, C. F. (2004), "Measuring Expectations", *Econometrica*, **72**, 1329–1376.
- MANSKI, C. F. (2006), "Interpreting the Predictions of Prediction Markets", *Economics Letters*, **91**, 425–429.
- MILGROM, P. R. and STOKEY, N. (1982), "Information, Trade and Common Knowledge", *Journal of Economic Theory*, **26**, 17–27.
- MITCHELL, D. (1989), *Winning Thoroughbred Strategies* (New York: W. Morrow).
- OTTAVIANI, M. and SORENSEN, P. N. (2006), "The Timing of Parimutuel Bets" (Mimeo).
- OTTAVIANI, M. and SORENSEN, P. N. (2010), "Noise, Information, and the Favorite-Longshot Bias in Parimutuel Predictions", *American Economic Journal: Microeconomics*, **2**, 58–85.
- OTTAVIANI, M. and SORENSEN, P. N. (2012), "Aggregation of Information and Beliefs: Asset Pricing Lessons from Prediction Markets" (Mimeo).
- POTTERS, J. and WIT, J. (1996), "Bets and Bids: Favorite-Longshot Bias and Winner's Curse" (Mimeo).
- QUANDT, R. E. (1986), "Betting and Equilibrium", *Quarterly Journal of Economics*, **101**, 201–207.
- SADKA, R. and SCHERBINA, A. (2007), "Analyst Disagreement, Mispricing, and Liquidity", *The Journal of Finance*, **62**, 2367–2403.
- SANDRONI, A. and SMORODINSKY, R. (1999) "The Speed of Rational Learning", *International Journal of Game Theory*, **28**, 199–200.
- SAUER, R. D. (1998), "The Economics of Wagering Markets", *Journal of Economic Literature*, **36**, 2021–2064.
- SERRANO-PADIAL, R. (2012), "Naive Traders and Mispricing in Prediction Markets", *Journal of Economic Theory*, **147**, 1882–1912.
- SHI, Z., ZHENG, T. and HAN, J. (2012), "Identifiability of Multivariate Logistic Mixture Models", *ArXiv e-prints*.
- SHIN, H. S. (1991), "Optimal Betting Odds Against Insider Traders", *Economic Journal*, **101**, 1179–1185.
- SHIN, H. S. (1992), "Prices of State Contingent Claims with Insider Traders and the Favourite Longshot Bias", *Economic Journal*, **102**, 426–435.
- SNOWBERG, E. and WOLFERS, J. (2010), "Explaining the Favorite-Longshot Bias: Is it Risk-Love, or Misperceptions?", *Journal of Political Economy*, **118**, 723–746.
- STEIN, J. C. (2009), "Presidential Address: Sophisticated Investors and Market Efficiency", *The Journal of Finance*, **64**, 1517–1548.
- THALER, R. H. and ZIEMBA, W. T. (1988), "Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries", *Journal of Economic Perspectives*, **2**, 161–174.
- TOMPKINS, R. G., ZIEMBA, W. T. and HODGES, S. D. (2008), "The Favorite-Longshot Bias in S&P 500 and FTSE 100 Index Futures Options: The Return to Bets and the Cost of Insurance", in Hausch, D. B. (ed), *Handbook of Sports and Lottery Markets* (Amsterdam, Netherlands: Elsevier B. V.) 161–178.
- TOWNSEND, R. M. (1978), "Market Anticipations, Rational Expectations, and Bayesian Analysis", *International Economic Review*, **19**, 481–494.
- TRAIN, K. E. (2003), *Discrete Choice Methods with Simulation* (New York, NY: Cambridge University Press).
- TVERSKY, A. and KAHNEMAN, D. (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty", *Journal of Risk and Uncertainty*, **5**, 297–323.
- VIVES, X. (1993), "How Fast do Rational Agents Learn?", *The Review of Economic Studies*, **60**, 329–347.
- VUONG, Q. H. (1989), "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses", *Econometrica*, **57**, 307–333.
- WHITE, H. (1982), "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, **50**, 1–25.
- WHITE, H. (1987), "Specification Testing in Dynamic Models", in BEWLEY, T. F. (ed), *Advances in Econometrics, Fifth world congress* (Cambridge: Cambridge Press), 1–58.