



香港城市大學
City University of Hong Kong

專業 創新 胸懷全球
Professional · Creative
For The World

CityU Scholars

Modeling fixed odds betting for future event prediction

Chen, Weiyun; Li, Xin; Zeng, Daniel

Published in:

MIS Quarterly: Management Information Systems

Published: 01/06/2017

Document Version:

Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

Publication record in CityU Scholars:

[Go to record](#)

Published version (DOI):

[10.25300/MISQ/2017/41.2.14](https://doi.org/10.25300/MISQ/2017/41.2.14)

Publication details:

Chen, W., Li, X., & Zeng, D. (2017). Modeling fixed odds betting for future event prediction. *MIS Quarterly: Management Information Systems*, 41(2), 645-665. <https://doi.org/10.25300/MISQ/2017/41.2.14>

Citing this paper

Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights

Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission

Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy

Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.

MODELING FIXED ODDS BETTING FOR FUTURE EVENT PREDICTION¹

Weiyun Chen

Department of Educational Information Technology, Faculty of Education, East China Normal University,
Shanghai, CHINA {weiyun.chen@qq.com}

Xin Li

Department of Information Systems, College of Business, City University of Hong Kong,
Hong Kong, CHINA {xin.li.phd@gmail.com}

Daniel Zeng

Department of Management Information Systems, University of Arizona, Tucson, AZ 85721 U.S.A., and
State Key Laboratory of Management and Control for Complex Systems, Institute of Automation,
Chinese Academy of Sciences, Beijing, CHINA {zeng@email.arizona.edu}

Prediction markets provide a promising approach for future event prediction. Most existing prediction market approaches are based on auction mechanisms. Despite their theoretical appeal and success in various application settings, these mechanisms suffer from several major drawbacks. First, opinions from experts and amateurs are treated equally. Second, continuous attention from participants is assumed. Third, such mechanisms are subject to various forms of market manipulation. To alleviate these limitations, we propose to employ the classic fixed odds betting as an alternative prediction market mechanism. We build a structural model based on a belief–decision framework as the event probability estimator. This belief–decision framework models bettors’ beliefs with mixed beta distributions and bettors’ decisions with prospect theory. A maximum likelihood approach is applied to estimate the model parameters. We conducted experiments on three real-world betting datasets to evaluate our proposed approach. Experimental results show that fixed odds betting based prediction outperforms the reduced form models based on odds and betting results, and achieves a comparable performance with auction-based prediction markets. The results suggest the possibility of employing fixed odds betting as a prediction market in a variety of application contexts where the assumptions made by auction-based approaches do not hold.

Keywords: Prediction market, fixed odds betting, crowd intelligence, prospect theory, decision support

Introduction

Fixed odds betting is a simple and time-tested gambling mechanism widely used in sports gambling (Song et al. 2007).

¹Kartik Hosanagar was the accepting senior editor for this paper. Ramesh Shankar served as the associate editor.

The appendix for this paper is located in the “Online Supplements” section of the *MIS Quarterly*’s website (<http://www.misq.org>).

In fixed odds betting, bookmakers set a constant payoff on event outcomes (i.e., odds) and call for bettors’ participation. Since contract payoff is critical to bookmakers’ financial gains, they often make an effort to employ domain experts and prediction models to predict the outcome when setting the odds. Bettors decide which side to bet based on the offered odds and their belief about the outcome. When the betting game finishes and the event outcome is known, the bettors who bet correctly will get paid. The central thesis of our research is to explore the possibility of using fixed odds betting as a prediction market mechanism.

Prediction markets are future markets based on events' outcomes (Wolfers and Zitzewitz 2004). They have been successfully used in areas such as presidential elections (Forsythe et al. 1992), movie box office numbers (Gruca et al. 2003), and infectious disease surveillance (Polgreen et al. 2007). In prediction markets, contract payoffs are associated with the outcomes of future events. Such markets encourage participants to reveal their private information in predicting future events. Prediction markets aggregate such information and are able to provide overall predictions that are often more accurate than traditional prediction models (Arrow et al. 2008). Participants in prediction markets are most concerned about their financial benefit from trading on the market. The prediction market organizer (i.e., market maker) is more interested in the prediction results, which can be used to make future policy, business, or investment decisions. In this kind of application setting, a prediction market must close before the real event occurs to allow market makers to take action. Traditional prediction markets are often established based on various auction formats, such as continuous double auction (CDA) (Wolfers and Zitzewitz 2004). CDA is often combined with an automated market maker using the logarithmic market scoring rule (LMSR) (Hanson 2003; Healy et al. 2010) or the dynamic pari-mutuel (DPM) method (Pennock 2004) to address some of its issues, such as the thin market problem. There is extensive literature on these auction-based prediction market mechanisms.

Compared with existing prediction markets, fixed odds betting has some interesting features. First, it distinguishes expert opinions from nonexpert opinions. Ignoring the impact of expert opinions on other participants' decisions (Renn et al. 1993) was considered as one of the main reasons for the failure of terrorism prediction markets (Hanson 2006a, 2006b; Meirowitz and Tucker 2004). In fixed odds betting, to some extent, odds reflect expert opinions and bets reflect crowd intelligence. This distinction allows the possibility of modeling experts' and participants' opinions separately and studying their interactions. Second, fixed odds betting requires participants choose which side to bet on, which requires much less cognitive effort than auctions impose. Auction-based mechanisms usually demand participants specify their private information in the form of contract price (Madhavan 2012), which puts a high cognitive load (Parkes et al. 1999) on participants and limits the number of participants (Graefe and Armstrong 2011; Green et al. 2007). Third, auction-based mechanisms are subject to price-based market manipulation due to the use of dynamic contract prices (Hansen et al. 2004; Rothschild and Wolfers 2008). As in financial markets, dishonest traders can use their claimed price as a tool to influence the actions of other traders, distort prices, and profit from the marketplace (Blume et al. 2010). Such manipulative acts are not possible in fixed price betting.

Leveraging the crowd intelligence of bettors and using fixed odds betting as a prediction market leads to a unique challenge: fixed odds betting does not have an appropriate instrument to aggregate the crowd's private information. Prior studies have used odds to generate predictions (Kain and Logan 2014; Sauer 1998). However, in these studies, bookmakers' odds setups are based on proprietary models rather than participants' information. Since participants' bets are made under the influence of odds, Kain and Logan (2014) argue that they do not really predict event outcomes. To repurpose fixed odds betting as a prediction market mechanism, it is necessary to build an event probability estimator as an enabling module.

To model fixed odds betting and build the event probability estimator, we draw from the literature on decision-making under uncertainties and propose a structural model for individuals' betting decisions in a belief–decision (BD) framework. In this framework, we assume that participants hold private information (i.e., a belief) that is correlated with an event's outcome. Participants' beliefs are modeled by mixed Beta distributions parameterized on event probability. We then infer individuals' betting decisions under the influence of odds to maximize their expected utility, which is based on prospect theory. A maximum likelihood approach is used to estimate the parameters of this BD model. We conducted computational experiments on three real-world betting datasets involving sports and entertainment events to evaluate our proposed model. We found that our model can deliver comparable performance with auction-based prediction markets. The performance of our model also improves with the prediction power of the odds specified by the market maker. These findings demonstrate the feasibility of using fixed odds betting as an alternative prediction market, potentially enabling applications of prediction markets in situations where the standard auction-based mechanisms are not appropriate.

Our research is of both theoretical and practical value. Theoretically, understanding crowd decisions under external influence, manifested through the impact of odds, is of great interest to prediction market and crowd intelligence research. Fixed odds betting provides us with a unique opportunity to study this problem. Practically, our research demonstrates that it is possible to make use of widely available fixed odds betting platforms to make predictions that are of value to society, some of which may not suit auction-based prediction markets. As part of our study, we have developed an approach to estimate event probability along with the distribution of individuals' beliefs, which could be applied in various settings involving crowd intelligence.

The paper is structured as follows. The “Background” section reviews existing prediction market mechanisms and fixed-odds betting. The subsequent section elaborates our fixed odds betting based prediction market design and presents our proposed event probability estimator. We then validate our model, followed by a discussion of the implications and limitations of our proposed mechanism. Finally, we present our conclusions.

Background

Prediction Markets

Early forms of prediction markets appeared in 1884 to predict the presidential election (Rhode and Strumpf 2008). Recent years have seen a surge in electronic prediction markets, such as the Iowa Electronic Markets (Berg and Rietz 2006) and Intrade.com (Erikson and Wlezien 2008). Prediction markets have a high accuracy in predicting future events (Arrow et al. 2008) and are being employed in predicting U.S. presidential elections (Forsythe et al. 1992), Hollywood movie box office numbers (Gruca et al. 2003), Swine Flu pandemics (Ritterman et al. 2009), and companies’ financial results (Bondarenko and Bossaerts 2000). They have also been proposed for terrorist analysis (Hanson 2006b), product design (Dahan et al. 2010), and supply chain management (Guo et al. 2006).

Trading mechanisms are the core of prediction markets. They affect the participants’ behavior and market efficiency (Chen et al. 2010; Jian and Sami 2012). Auctions are the most widely used trading mechanisms, successfully supporting several prediction markets worldwide.

Continuous double auction (CDA) is a classic prediction market trading mechanism (Wolfers and Zitzewitz 2004). In a CDA market, a participant can submit a buy/sell request that specifies intended price and number of contracts. Orders are matched according to price and request time, whereas bid and ask price changes in the market reflect traders’ digestion of new information related to the event (Lee and Moretti 2009). Since CDA leads to a zero-sum game, it brings no financial risk to the market maker itself. However, it requires both interested sellers and buyers. If there are not enough participants, the thin market problem arises (Hanson 2003; Pennock 2004), in which some participants’ transaction requests cannot be fulfilled and their information cannot be incorporated into the prediction.

Auction-based mechanisms usually assume that participants know their private price on goods (Madhavan 2012) and can

actively track contract price changes and respond accordingly. This high and continuous cognitive load (Parkes et al. 1999) limits the number of participants (Graefe and Armstrong 2011; Green et al. 2007). To address the thin market problem, one can either set up a market maker (i.e., an artificial trader) to trade with other participants or reduce the barrier for participants to use the market (Pennock 2004).

The logarithmic market scoring rule (LMSR) introduces an automatic market maker to alleviate the thin market problem (Hanson 2003). It automatically adjusts bid/ask price based on contract inventory. LMSR increases the liquidity of the market and guarantees that the market maker’s loss is within some boundaries. However, under LMSR, participants need to mentally determine the contract price by considering the nonlinear relationship between price and trading volume (Blohm et al. 2011), which requires even higher cognitive load on the side of the participants.

The dynamic pari-mutuel mechanism (DPM) is another proposed solution to the thin market problem, combining CDA with pari-mutuel betting (i.e., winner-takes-all) for prediction markets (Pennock 2004). In pari-mutuel betting (Quandt 1986), the bidders choosing the correct side will get and share the money bid by all bidders. Such a simple betting rule can function under a thin market, which also leads to final payoff odds reflecting event probability (Asch et al. 1982) with a favorite-longshot bias (Quandt 1986). However, in pari-mutuel betting, later participants have a competitive advantage due to having more accurate payoff information (Asch et al. 1982; Pennock 2004), which discourages participants from joining the market in a timely manner (Pennock 2004). To avoid this problem, DPM has an automated market maker sell the contract at a price determined by demand. Such contracts are associated with a payoff determined when the market closes. While addressing the waiting problem, DPM introduces nonlinearity into the pricing function, which increases the participants’ cognitive load (Blohm et al. 2011).

Although the thin market problem is partially alleviated through innovations in trading mechanisms, the limited number of participants still affects the power and applicability of prediction markets (Healy et al. 2010). In addition, auction-based mechanisms do not distinguish experts from regular participants (Abramowicz and Henderson 2007; Graefe and Armstrong 2011). Utilizing crowd intelligence may be more difficult in circumstances where expert opinions are influential (Onkal et al. 2009). Finally, auction-based mechanisms are subject to price-based market manipulation (Hansen et al. 2004; Rothschild and Wolfers 2008), which could further limit the usefulness of prediction markets.

Fixed Odds Betting

Fixed odds betting is a time-tested mechanism widely used in the betting market (e.g., in UK football game betting). Bookmakers (i.e., market makers) set the payoff (odds) before the betting starts and usually do not change it during the betting period. Bettors (i.e., participants) decide which side to bet based on the offered odds. When the betting game finishes, the bettors are rewarded with the specified payoff (per bet) if their betting is correct.

Previous fixed odds betting research has primarily focused on the bookmakers' perspective, studying problem such as odds setup and the efficiency of fixed odds betting.

Odds setup is a critical problem in fixed odds betting. If the odds are mispriced and many bettors make correct predictions with high payoffs, the bookmaker will face significant economic loss (Kuypers 2000). As such, bookmakers usually make an effort to predict future event outcomes and tweak the odds to balance the amount of betting. Bookmakers' odds setups are typically based on proprietary models or expert opinions, which shows strong positive correlation with actual event outcomes (Song et al. 2007). In recent studies, odds are found to be increasingly effective in predicting future events (Forrest et al. 2005; Strumbelj and Sikojca 2010), possibly indicating more significant analytics efforts by bookmakers. Spann and Skiera (2009) found that betting odds outperforms independent experts in predicting future events, and there have been attempts to make use of betting odds for predictions (Kain and Logan 2014; Sauer 1998).

The second major question is whether fixed odds betting is an efficient market (i.e., whether the odds reflect the information from the market). Gabriel and Marsden (1990) found that the British racetrack betting market does not meet the condition of semi-strong efficiency. Goddard and Asimakopoulos (2004) found that English league football betting is weak-form inefficient if historical match data and other explanatory variables are considered. Kuypers (2000) shows that it is possible for bookmakers to take advantage of bettors' behavioral bias (Woodland and Woodland 1994), set inefficient market odds, and make a profit. Steven (2004) argued that bookmakers are better at predicting game outcomes than typical bettors and thus can systematically exploit bettors' bias by choosing appropriate odds.

Although the odds have been shown to correlate with event probability, they do not capture participants' private information. They only reflect market makers' information and analysis. Kain and Logan (2014) even argued that participants' bets do not predict event outcomes and cannot be directly used for prediction. It is clear that work enabling the

use of participants' information in predicting future events is lacking in the current literature.

Decisions Under Uncertainty

To exploit crowd intelligence in fixed odds betting, it is necessary to model the bettor's decision process. In horse racing betting, Jullien and Salanie (2000) modeled participants' attitudes toward risk using three types of models—rank-dependent utility, expected utility, and cumulative prospect theory—and found that the prospect theory model has strong explanatory power. In previous research, Chen et al. (2011) took a simulation approach to derive an approximated reduced form model on the relationship between odds, the ratio of bettors' choices, and bettors' average belief.

In related fields, people's decisions under uncertainty are also being studied extensively. For example, in the marketing literature, Erdem and Keane (1996) studied consumers' brand choices under the influence of their private information and advertisement exposure. They developed a structural model that assumes consumers made decisions to maximize their utility, which significantly outperforms a reduced form model. In social media research, Dai et al. (2012) modeled users' private information, inter-user influence, and restaurant quality change in online ratings, aiming to reveal the more objective ratings. Through a structural model on users' evaluations of products based on their experience and product reviews, Zhao et al. (2013) found that online reviews are more important than user experience in affecting purchase decisions. It is argued that structural models have a better theoretical basis and less chance of making misleading forecasts than reduced form models (Chintagunta et al. 2006).

Fixed Odds Betting-Based Prediction Market

Although betting results and odds are not a natural prediction market, fixed odds betting possesses several desirable characteristics, including the separation of expert opinions and participants' decisions, low cognitive load, and low risk of market manipulation. In our study, we attempt to develop a novel prediction market mechanism based on fixed odds betting.

Market Setup

Figure 1 shows the overall design of our proposed fixed odds betting-based prediction market. This approach has two major

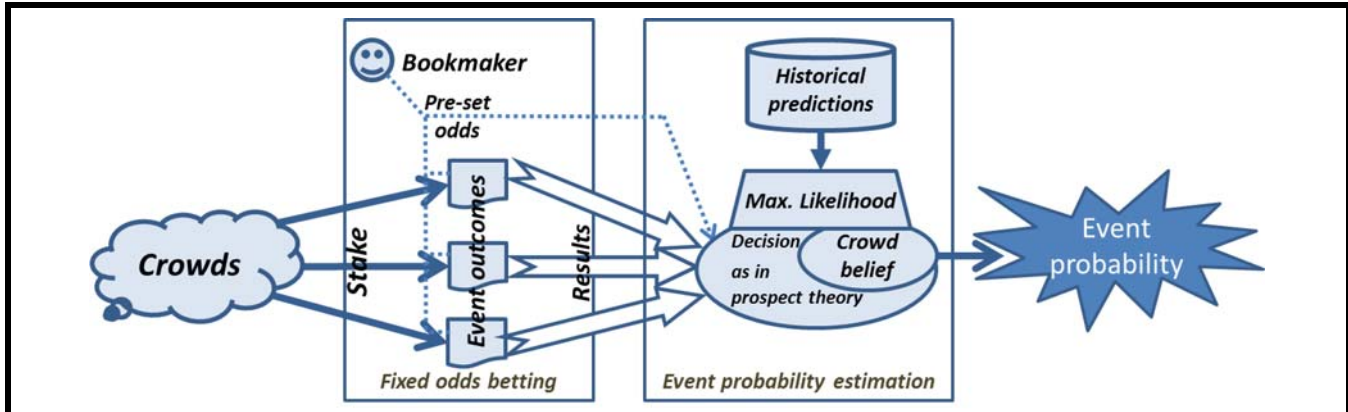


Figure 1. Fixed Odds Betting-Based Prediction Market

components: betting and event probability estimation. The betting component is a standard fixed odds betting module with a market maker/bookmaker setting up fixed odds on event outcomes and participants staking on their preferred event outcomes. To repurpose fixed odds betting as a prediction market mechanism, it is necessary to have an event probability estimator to “decode” event probability from the betting processes and results. Our work on this event probability estimation module is intended as a major technical contribution of the reported research.

Without loss of generality, we consider predicting events with binary outcomes, say, A and B . When appropriate, we use A or B in subscript to illustrate whether a variable is measured on A or B . The (hidden) probabilities for two outcomes are p_A and p_B , respectively, where $p_A + p_B = 1$. In the fixed odds setup, the bookmaker needs to specify odds for outcomes, o_A and o_B , where $o_A, o_B > 0$. If a participant’s betted outcome occurs, she can get the original ante back plus an extra payoff equal to the odds times ante (Kuypers 2000). Here we assume “no choice” is not a choice since such participants have already been self-selected and are not observable to the bookmaker. When the betting finishes, we will observe the total number of participants m and the percentages of participants (i.e., bet ratio) on each side, s_A and s_B , where $s_A + s_B = 1$. The actual event outcome is R , which is 1 for A and 0 for B . Since the events have binary outcomes, this paper focuses on the dynamics from the point of view of A . The changes from B ’s point of view can be equally derived. For simplicity, we remove the subscript A if it causes no ambiguity.

Although the technology to implement the betting part of the prediction market is the same as that for fixed odds betting, its different purposes bring out some subtle changes in the setup. Specifically, prediction market organizers are concerned with capturing crowds’ intelligence while fixed odds betting book-

makers are concerned with obtaining financial gain from the market. As a result, the objective for odds setup and the timing of market operations are changed.

In fixed odds betting, the odds setup problem has been thoroughly discussed from the perspective of avoiding financial loss (Kuypers 2000). To avoid participants systematically gaining from betting on both sides, bookmakers need to set $d = 1/(1 + o_A) + 1/(1 + o_B)$ within range $[1,2)$. Parameter d is called over-roundness and represents the profit level of the bookmaker. In sports betting, d is often set to 1.11 (Kuypers 2000). After determining d , bookmakers often need to hire domain experts and build proprietary models to estimate event probability on A (or ratio of bettors holding different opinions), noted as p_{odds} hereafter. From the perspective of avoiding financial loss, the bookmaker’s odds setup should balance the bets on the two sides so that no side has a higher expected utility that can be taken advantage of by the bettors. For our binary event outcome setup, this means $(o_B + 1)/[(o_A + 1) + (o_B + 1)] = p_{odds}$ (Kuypers 2000).² There are infinite pairs of o_A and o_B that can fulfill this constraint. However, given a predetermined d , the odds are set as

$$o_A = \frac{1}{p_{odds}d} - 1; \quad o_B = \frac{1}{(1 - p_{odds})d} - 1^3$$

²The implied probability of outcome A p_A is proportional to $1/(1 + o_A)$. From the bookmaker’s perspective, under such a probability, the expected return of loss (the \$1 bid) is equal to the expected return of gain (the o_A payoff). Regardless of the outcome, the bookmaker, statistically, will not lose money. In our context, there are two outcomes, and the sum of their probability is $p_A + p_B = 1$. The probability of event $p_{odds} = p_A$ can be calculated through a normalization procedural: $1/(o_A + 1)/[1/(o_A + 1) + 1/(o_B + 1)]$, which is equal to $(o_B + 1)/[(o_A + 1) + (o_B + 1)]$.

³Since $d = 1/(o_A + 1) + 1/(o_B + 1)$, $p_{odds} * d = 1/(o_A + 1)$. Therefore, $o_A = 1/(p_{odds} * d) - 1$.

From the prediction market's perspective, when using the above formulae to set odds, market makers want to select a smaller d (i.e., make the odds higher) to attract participants to contribute their private information. If financial loss on the betting market is a major concern, market makers need to solve an optimization problem considering the joint effects of profit level (for a short-term return from the betting) and prediction accuracy (for a long-term return from the event). To avoid such complexity in market design, the prediction market can operate with virtual money and reward participants based on the ranking of their accumulated virtual money. This is a common practice to bound rewards (Wolfers and Zitzewitz 2006) and avoid legal issues. The virtual money market performs as well as the real money market (Servan-Schreiber et al. 2004) and provides the market maker with more flexibility for setting odds.⁴

In the betting markets on score-based games, such as football matches, bookmakers may set up a "spread," which refers to the point difference between the two sides. The purpose is to balance bettors and make the betting game more attractive. The spread is correlated with the outcome of the match (Kain and Logan 2014). However, setting a spread changes the event prediction. The spread becomes a part of the revised event (e.g., team A will beat team B by more than five points) and cannot be used to predict the revised event. In a prediction market, the spread cannot replace the role of odds in our framework. In addition, the spread represents the market maker's knowledge (as do the odds), which is different from our objective of leveraging crowd intelligence to improve prediction.

The incentive of the market maker is to obtain economic gain outside of the prediction market by leveraging the crowd's intelligence. The participants are incentivized by financial benefits or social rankings received from the market. Since the market uses financial- or reputation-based incentives to motivate participants to reveal their private information (Schreiber 2009), we argue that our method is superior to simple information aggregation methods, such as polling. Besides, to work as a prediction market, the betting game needs to end before the event's occurrence, so that the market maker has time to extract information from the prediction market for prediction.

⁴This research does not deal with the problem of odds setup considering both short-term and long-term return. In our model, we consider odds to be exogenous. In our experiments, the datasets are operated with virtual money. Thus, the bookmaker's odds setup will not be affected by their short-term return in the betting market. We do not have clear evidence on what bookmakers do to estimate event probability and set up the odds.

An Illustrative Example

To better illustrate our approach, in this section we introduce a hypothetical example. Assume, for the sake of strategic planning, an organization is interested in predicting which political party (either the Republican or Democratic) will win the next U.S. presidential election. This organization acquires predictions from experts. In addition, recognizing that ordinary citizens' opinions play an important role in the electoral process and their votes will eventually decide the outcome of this election, this organization has set up a prediction market.

Assuming the experts' predicted winning probability for a Republican candidate is 40% and the organization, which functions as a bookmaker, sets d as 1.11, the odds would be 1.25 for betting on the Republican candidate and 0.5 for betting on the Democratic candidate, according to the aforementioned odds setup formulae. If a market participant knows exactly who will win the next presidential election, she will not be affected by the odds. For participants who do not know the future, their choice would depend on their estimated probabilities that the Republican candidate will win.⁵ For example, if a person's estimated winning probability for the Republican candidate is 49%, her choices would be as follows:

- A. Betting on Republicans: "49% chance to win \$1.25," versus
- B. Betting on Democrats: "51% chance to win \$0.5."

Her choice is affected by her risk-taking profile. For instance, she might bet on the Republican candidate to trade the small difference in winning probability for the large difference in return.

At the crowd level, the number of participants making betting decisions depends on their belief distribution and the shape of their valuation function on monetary returns. For example, if 30% of participants consider the Republican candidate's winning probability to be 49% and the others have a firm belief that their side will win, then these 30% face the above binary choice. If we know *a priori* the valuation function of these participants, we can project their betting decision.

⁵Note that users may not explicitly consider the probability and do math as we do. They directly reason in their minds and provide the binary judgement.

As we can see, in this hypothetical context, the final betting (ratio) is jointly influenced by individual participants' knowledge and the bookmaker's knowledge. Our proposed approach untangles the interactions between the two types of knowledge. A critical component of our proposed model is how to estimate participants' belief distribution.

A Belief–Decision Framework for Event Probability Estimation

We propose a structural model, which models individuals' betting decisions to estimate the crowd's belief and decisions. This model is referred to as the belief–decision framework. In comparison with a reduced form model (Chen et al. 2011), the structural model has a stronger theoretical basis that incorporates individuals' reasoning and decisions.

Consider participant r who bets on event i with event probability p_i within $(0, 1)$.⁶ (For simplicity, we remove the subscript r and i below if there is no ambiguity.) She knows her belief b_{ri} and observes odds o_i to make a bet. Following the previous studies (Dai et al. 2012; Erdem and Keane 1996), we assume that the participant's belief b_{ri} is a random variable following an *i.i.d.* distribution. We start with a set of necessary conditions on the distribution of belief b_{ri} , considering the nature of a prediction market. In the next subsection, we specify $b_{ri} \sim f(x; p_i, \theta)$ as a mixed Beta distribution depending on p_i and some parameters θ , which meets our modeling requirements as proved in Theorems 1 and 2.

Given odds, rational participants make betting decisions to maximize their expected utility. In the subsection on individuals betting decisions, we leverage prospect theory to quantify participants' responses to monetary returns and determine their choices. In this process, we assume that the participants are homogenous as to their utility functions, risk attributes, and decision processes. The underlying assumption is that the differences among these participants are reflected through differences in held belief, which in turn results in different betting decisions. Note that we do not make any assumptions about odds, as it is assumed to be exogenous, as determined by the market maker.

Since parameters θ in our model are not known *a priori*, we learn such parameters from historical data. In the subsection on parameter estimation, we apply a maximum likelihood approach to estimate parameters θ . Theorems 3 and 4 show the validity of this approach. Theorem 5 ensures that we can

⁶We exclude the values 0 and 1 for mathematic convenience in the model.

infer a unique p_i from the data given our assumptions. We further assume each participant makes only one bet. Participants with multiple bets are considered different participants.

In the final subsection, we illustrate this process using the hypothetical presidential election example.

Private Information/Belief Distribution

For participant r and event i , we assume that her belief distribution $f(x; p, \theta)$ follows three conditions:

Condition 1. $\int_0^1 f(x; p, \theta) = 1$, that is, the participant's private information on event probability is within $[0, 1]$.

Condition 2. $f_A(x; p, \theta) = f_B(x; p, \theta)$. Note that A and B are arbitrary labels for the outcomes. The private information distributions should not change if we simply relabel them. Holding a belief x on outcome A also means holding a belief of $1-x$ on outcome B , that is, $f_B(x; p, \theta) = f_A(1-x; 1-p, \theta)$. Thus, $f(x; p, \theta) = f_A(x; p, \theta) = f_A(1-x; 1-p, \theta) = f(1-x; 1-p, \theta)$, that is, the information distributions on A and B are symmetric to line $x = 0.5$.

Condition 3. $E[f(x; p, \theta)]$ is strictly increasing with p . This allows the participant's belief to have a positive correlation with the event's probability. That is, the crowd opinions have a predictive power on the event, which is the basic premise behind prediction markets.

Given these three conditions, we assume that the belief distribution follows a mixture of independent Beta distributions. Beta distribution is a flexible density function, and its mixtures can approximate the shape of a large range of probability density functions, especially for subjective belief (Audun 1997). We assume that the belief distribution $f(x; p, \theta)$ is composed of D Beta distribution components:

$$f(x; p, \theta) = \sum_{l=1}^D \frac{\lambda_l}{\sum_{j=1}^D \lambda_j} \cdot \text{Beta}(x; \alpha_l(p), \beta_l(p)) \quad (1)$$

where λ_l is the positive coefficients of the mixed Beta distribution, and $\alpha_l(p)$ and $\beta_l(p)$ are shape parameters for Beta distribution. For our modeling purpose, we specify $\lambda_l > 0$, $\alpha_l(p) > 0$, and $\beta_l(p) > 0$; for $p \in (0, 1)$. Here, we normalize on λ_l to ensure the total probability integrates to unity.

When we specify the parameters of equation (1), $f(x; p, \theta)$ needs to meet the three conditions we proposed. Since the

Beta distribution is within range [0, 1] and the mixed Beta distribution is also within range [0,1], Condition 1 is automatically fulfilled. For the second condition, it can be proved that a sufficient condition guaranteeing Condition 2 is $\alpha_i(p) = \beta_i(1-p)$:

Theorem 1: If $\alpha_i(p) = \beta_i(1-p)$, then $f(x;p,\theta) = f(1-x;1-p,\theta)$. [Proof in Appendix.]

Since $f(x;p,\theta)$ is a linear combination of Beta distributions, Condition 3 is met if the mean of each Beta distribution component increases with p .

Considering the above requirements, we specify $\alpha_i(\cdot)$ and $\beta_i(\cdot)$ as polynomial functions:

$$\alpha_i(p) = \sum_{h=1}^Z u_{ih} p^h, \quad \beta_i(p) = \sum_{h=1}^Z u_{ih} (1-p)^h \quad (2)$$

where Z is the degree and u_{ih} is the nonnegative coefficient of polynomial components. Obviously, this specification satisfies Condition 2 since $\alpha_i(p) = \beta_i(1-p)$ and also Condition 3.

Theorem 2: $E[\text{Beta}(x; \alpha_i(p), \beta_i(p))]$ is strictly increasing with p . [Proof in Appendix.]

Note that $\lim_{p \rightarrow 0} E[\text{Beta}(x; \alpha_i(p), \beta_i(p))] = 0$, $\lim_{p \rightarrow 1} E[\text{Beta}(x; \alpha_i(p), \beta_i(p))] = 1$, and if $p = 0.5$, $E[\text{Beta}(x; \alpha_i(p), \beta_i(p))] = 0.5$. Thus, $E[f(x;p,\theta)]$ increases from 0 to 1 when p increases from 0 to 1. The mean of the participant's belief has a positive correlation with the event probability.

In equations (1) and (2), we specify the plausible parametric form of crowd belief distribution with parameter

$$\theta \in \Theta = \{ \lambda_l > 0; u_{ih} > 0 \mid l = 1, \dots, D; h = 1, \dots, Z \}$$

Individuals' Betting Decisions

Given a participant's belief and information concerning the odds, a utility-maximizing participant will bet on the side that provides the higher expected utility. Following the literature, we model individual decision behavior with cumulative prospect theory (Tversky and Kahneman 1992), which specifies the utility function $U(\cdot)$ given belief b and odds o as

$$U(b, o) = w^+(b)v(o) + w^-(1-b)v(-1) \quad (3)$$

where $v(\cdot)$ is the valuation function for gains and losses, and $w^+(b)$ and $w^-(1-b)$ are weighting functions of the belief. When betting a monetary unit, the probable gain is o and probable loss is 1. Their valuations are $v(o)$ and $v(-1)$, respectively. Fully understanding participants' utility function in prediction markets requires extra experiments. In our research, we chose the following commonly used functional form from cumulative prospect theory:

$$w^+(x) = x^\gamma / [x^\gamma + (1-x)^\gamma]^{1/\gamma}, \quad w^-(x) = x^\tau / [x^\tau + (1-x)^\tau]^{1/\tau} \quad (4)$$

$$v(y) = \begin{cases} y^q & \text{if } y \geq 0 \\ -\rho(-y)^p & \text{if } y < 0 \end{cases} \quad (5)$$

and set q and p as 0.88, ρ as 2.25, γ as 0.61, and τ as 0.69, following the previous literature (Tversky and Kahneman 1992). The original experimental scenario for which these parameter settings were developed is very similar to our application scenario. Field estimation results also suggest that cumulative prospect theory fits the observed fixed odds betting data quite well (Jullien and Salanie 2000).

With $U(\cdot)$ specified, in our binary prediction scenario, if the utility of outcome A , $U(b, o_A)$, is larger than that of outcome B , $U(1-b, o_B)$, a participant will bet on A , and vice versa. It can be proved that if parameters $q, p, \rho, \gamma, \tau, o_A$, and o_B are positive, there exists only one belief value $c \in (0,1)$ providing equal utility to A and B , which is called balance belief hereafter.

Theorem 3: Given $U(x,o)$ specified by equations 3 to 5, if parameters $q, p, \rho, \gamma, \tau, o_A$, and o_B are positive, there exists only one belief value $c \in (0,1)$ s.t. $U(c, o_A) = U(1-c, o_B)$. [Proof in Appendix.]

Since the utility function is a continuous increasing function of belief, everyone who has a belief higher than c will bet on A . With the parameters in the utility function predetermined, balance belief can be solved from $U(c, o_A) = U(1-c, o_B)$ by providing o_A and o_B (using a standard root-finding algorithm, if necessary). Given the belief distribution $f(x;p,\theta)$ and the calculated balance belief c , we can calculate the probability a participant will bet on A as follows:

$$\begin{aligned}
 PA(p, \theta) &= P(X > c; p, \theta) = \\
 \int_c^1 f(x; p, \theta) dx &= \sum_{i=1}^D \frac{\lambda_i}{\sum_{j=1}^D \lambda_j} \cdot \int_c^1 Beta(x; \alpha_i(p), \beta_i(p)) dx \\
 &= \sum_{i=1}^D \frac{\lambda_i}{\sum_{k=1}^D \lambda_k} \cdot \left[1 - \int_0^c Beta(x; \alpha_i(p), \beta_i(p)) dx \right] \\
 &= \sum_{i=1}^D \frac{\lambda_i}{\sum_{k=1}^D \lambda_k} \cdot [1 - I_c(\alpha_i(p), \beta_i(p))]
 \end{aligned} \tag{6}$$

where $I_c(\alpha_i(p), \beta_i(p))$ is the regularized incomplete Beta function.

Parameter Estimation Based on Collective Behavior

With individual decision probabilities estimated by $PA(p, \theta)$, we now take a maximum likelihood approach to estimate the parameters provided that historical data are available. Suppose we have H binary fixed odds betting events, with observable data $\{o_{iA}, o_{iB}, R_i, m_i, s_{iA}, s_{iB}\}$ on odds, real outcome, total bets, and net bet ratio for each event i . Assuming that all bets are independent, the log likelihood function of observing the historical data is calculated as follows:

$$\begin{aligned}
 Lc(p, \theta) &= \log \left(\prod_{i=1}^H PA(p_i, \theta)^{m_i s_{iA}} (1 - PA(p_i, \theta))^{m_i (1 - s_{iA})} p_i^{R_i} (1 - p_i)^{1 - R_i} \right) \\
 &= \sum_{i=1}^H (m_i s_{iA} \log(PA(p_i, \theta)) + m_i (1 - s_{iA}) \log(1 - PA(p_i, \theta))) \\
 &\quad + \sum_{i=1}^H (R_i \log(p_i) + (1 - R_i) \log(1 - p_i))
 \end{aligned} \tag{7}$$

It can be proved that when m_i are sufficiently large, maximizing equation (7) will lead to $PA(p_i, \theta) = s_{iA}$.

Theorem 4: When m_i are sufficiently large, maximizing equation (7) will lead to $PA(p_i, \theta) = s_{iA}$. [Proof in Appendix.]

Computationally, following Theorem 4, given sufficiently large values for m_i , we first solve p_i from $PA(p_i, \theta) = s_{iA}$ and then estimate other parameters from the event outcomes. This is equivalent to maximizing the likelihood function on individual bets and event outcomes separately in Equation (7). To understand this solution from a different perspective, we could consider the betting process as a Bernoulli process since participants are assumed to be independent. According to

Casella and Berger (2002), the final observed bet ratio s_{iA} for the Bernoulli process i is an unbiased, asymptotically normal, and asymptotically efficient estimator of $PA(p_i, \theta)$ following the probability distribution:

$$\sqrt{m_i} (s_{iA} - PA(p_i, \theta)) \longrightarrow N(0, s_{iA} (1 - s_{iA})) \tag{8}$$

where m denotes the number of agents in the game. Thus, larger m_i leads to smaller estimation variance.

In short, as the first step of parameter estimation, we solve p_i from $PA(p_i, \theta) = s_{iA}$ as a function of o_A, o_B , and s_A , which is referred to as $p(o_{iA}, o_{iB}, s_{iA}; \theta)$. In Equation (6), $PA(p, \theta)$ is a linear combination of $1 - I_c(\alpha_i(p), \beta_i(p))$. Note that

Theorem 5: The regularized incomplete Beta function

$$I_c(\alpha_i(p), \beta_i(p)) = \int_0^c Beta(x; \alpha_i(p), \beta_i(p)) dx$$

is strictly decreasing with p . [Proof in Appendix.]

Thus, $PA(p, \theta)$ is strictly increasing with p . With all parameters specified, there always exists one and only one solution for $PA(p_i, \theta) = s_{iA}$. This solution can be found numerically.

After solving $p(o_{iA}, o_{iB}, s_{iA}; \theta)$, the second step of parameter estimation is to estimate $\theta \in \Theta$, by maximizing log likelihood function $Lc^*(\theta)$:

$$\begin{aligned}
 Lc^*(\theta) &= \log \left(\prod_{i=1}^H p(o_{iA}, o_{iB}, s_{iA}; \theta)^{R_i} (1 - p(o_{iA}, o_{iB}, s_{iA}; \theta))^{1 - R_i} \right) \\
 &= \sum_{i=1}^H (R_i \log(p(o_{iA}, o_{iB}, s_{iA}; \theta)) + (1 - R_i) \log(1 - p(o_{iA}, o_{iB}, s_{iA}; \theta)))
 \end{aligned} \tag{9}$$

$Lc^*(\theta)$ does not have a close-form solution. To find θ to maximize it, the classic gradient-based optimization cannot be used for technical reasons. In our approach, we employ the Nelder-Mead method (Kelley 1987) to solve this optimization problem [Figure A1 in the Appendix]. The Nelder-Mead algorithm maintains a set of testing points from the search space. To find the optimal solutions, the weakest ones in the set are replaced by applying Shrink, Expand, and Reflect operations according to the centroid of the set. These operations are equivalent to going closer to or farther from the centroid, or being reflected by the centroid, respectively. Although the Nelder-Mead algorithm is not always guaranteed to converge to optimal solutions (Kim 1998), its performance in practice is generally effective (Lagarias et al. 1998). In our experiments, we also randomly selected sets of starting points to reduce the chance of being trapped in local optima. Based on our computational experience, this algorithm generally converges quickly.

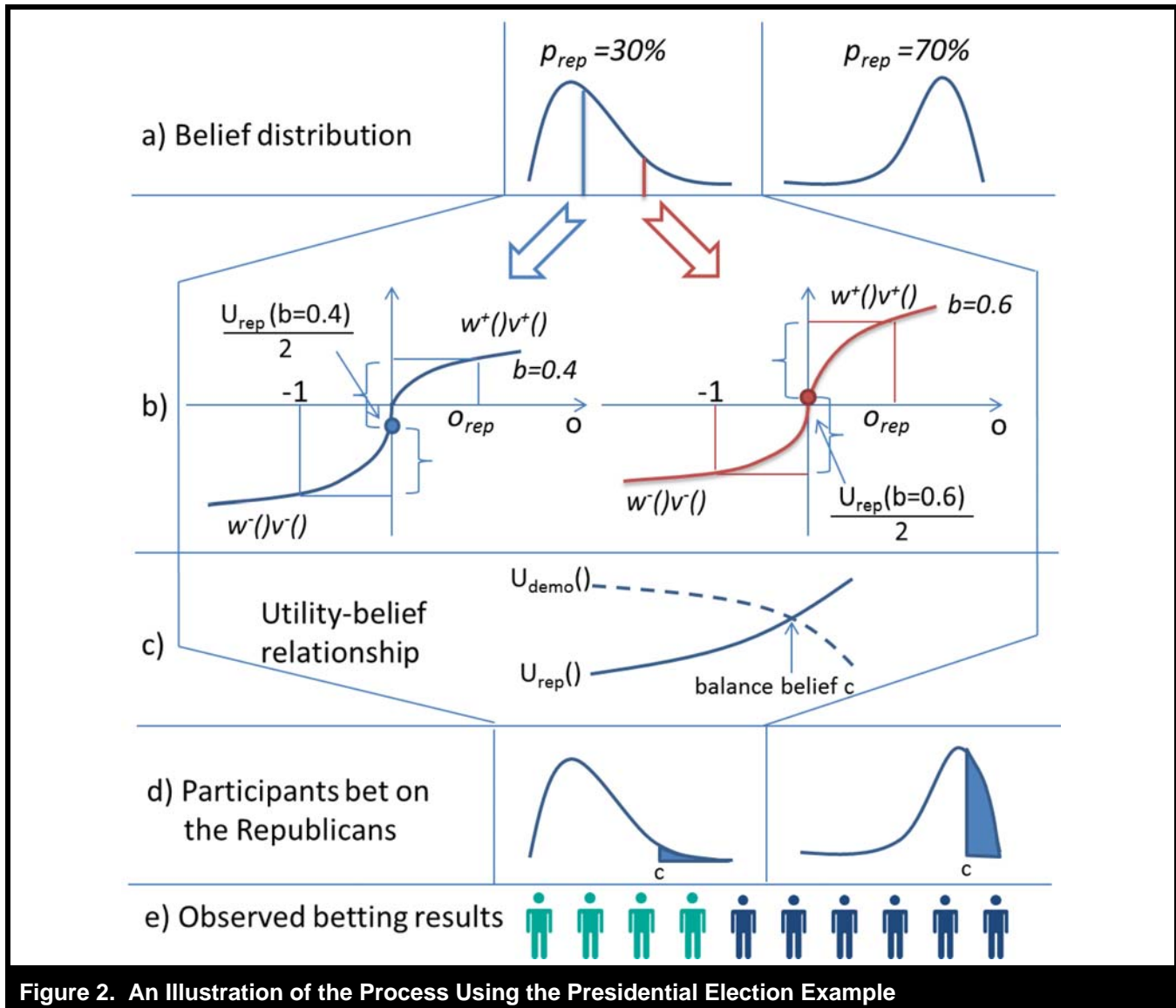


Figure 2. An Illustration of the Process Using the Presidential Election Example

It should be noted that parameter space θ depends on the ranges of D and Z . Larger D and Z lead to more complicated models, which may lead to the overfitting problem and reduce their generalizability. For computational tractability, we specified a range of D and Z before model selection. There are two potential approaches for model selection. The first is to choose the model with maximum likelihood. The second is to choose models based on AIC criteria (Burnham and Anderson 2002), where a smaller AIC is preferred. (Refer to the appendix for the AIC formula.) Following Burnham and Anderson (2002), we built composite models by combining models selected with AIC and weighting each component with $\exp((AICc_{min} - AICc_j)/2)$.

An Illustration of the Estimation Process

Figure 2 illustrates the above process with our hypothetical example of a presidential election. We assume the real and unknown winning probability for the Republicans is 30%. Most people have formed their belief on this event. The first subsection specified our assumed belief distribution function, which is bell-shaped and inclined to the side of the real winning probability as illustrated in Figure 2(a). As proved in Theorems 1 and 2, this specification meets the assumptions on the crowd's beliefs.

Figure 2(b) illustrates two people who hold the belief that the winning probability for the Republican candidate is 40% and

60%, respectively. To illustrate their decisions based on utility calculation, we show their weighted value function $w(\cdot)v(\cdot)$ as specified in the subsection on individuals' betting decisions, according to prospect theory. Given odds o_{rep} specified by the bookmaker, if one bets on the Republican candidate, the possible gain is o_{rep} and the possible loss is -1 for a \$1 bet. The mid point of the values of $w(\cdot)v(\cdot)$ on o_{rep} and -1 is half of the utility of betting on the Republican candidate. On the figure, the mid point for $b = 60\%$ is larger than that for $b = 40\%$. In fact, the weighted value function always passes (0,0) and depends on participants' beliefs. Theorem 3 proves that a higher belief b (on the Republican candidate) always leads to a higher valuation on the gain and loss. So, the utility of betting on the Republican (Democratic) candidate will increase (decrease) with the increase of belief b , as illustrated in Figure 2(c).

Since the relationship between utility and belief is a monotonic function, there exists a balance belief c , where people with a belief larger than c will bet on the Republican candidate. Such people are illustrated as the shadow region in Figure 2(d). If the model's parameters and the real winning probability are known, the above process will provide us with observed betting results that would match with the shadow region. If the model's parameters are known but the real winning probability is not known (e.g., to be chosen from $p_{rain} = 30\%$ or $p_{rain} = 70\%$), Theorem 5 ensures the feasibility and uniqueness of solving p from observed betting results.

If the model's parameters are unknown, as in the subsection on parametric estimation, we showed how to estimate these parameters from historical data using a maximum likelihood approach. In our hypothetical example, this is the same as holding pilot prediction markets on similar issues multiple times to get a sense of participants' belief and valuation functions.

Evaluation

Datasets

We evaluated our proposed BD framework using three real-world datasets. The first dataset is based on a fixed odds betting game organized by sina.com, one of China's largest online news websites, aiming to predict outcomes of a number of major contests as part of the 2008 Beijing Olympic Games. In this set of betting games, the website published odds 1 or 2 days before the contest began and stopped accepting stakes from participants about 1 or 2 hours before the contest finished. The betting was based on virtual currency, and the top 100 participants who accumulated the largest amount of

virtual money were awarded souvenirs. This set of sina.com betting games attracted more than 170,000 participants in total. We collected data on 167 distinct binary betting games; the numbers of participants in these game instances vary from 108 to 2,635. The mean number of participants per game is 564 with the standard deviation around 516. According to our observation, the betting ratio generally becomes stable when the betting finishes.

The second dataset used in our evaluation study is concerned with fixed odds games on entertainment events organized by sohu.com, another major news portal in China. The game setting was similar to that of sina.com but the odds setup was more *ad hoc*. From June 2008 to September 2014, sohu.com organized fixed odds game instances to predict 287 entertainment events, attracting more than 52,000 participants. From these events, we collected 34 binary betting game instances as our second dataset. For these 34 game instances, the numbers of participants varied from 5 to 1,964. The mean number of participants per game is 410 with standard deviation around 548.

The third dataset concerns 2014 FIFA World Cup betting games hosted on sohu.com. There were 278 such betting game instances, which attracted more than 112,000 participants. We collected information on 180 games with binary results. The numbers of participants for these binary games varied from 67 to 807. The mean number of participants per game is 331 with a standard deviation around 167.

Baseline Methods

For the purpose of future event prediction, one of the main baseline methods against which we compared our proposed model is Inklingmarkets.com, a leading auction-based prediction market. We first matched the events from our datasets with those from Inklingmarkets.com. Due to differences as to participants' interests and foci of the hosts of the prediction markets, the number of events for which predictions were available from both Inklingmarkets.com and the hosts covered in our datasets is relatively small. Ultimately, we were able to identify 6 events from the 2008 Beijing Olympic Games, 2 entertainment events, and 18 events from the 2014 FIFA World Cup. Since some of these identified game instances were not based on binary outcomes, we employed contract prices to represent event probabilities and computed relative probabilities of binary outcomes. These instances had the same spread setup on the game results as their betting counterparts, if available. The mean number of participants in these matched games is 62.

The second set of baseline methods is the bookmakers' estimation inferred from the odds/betting line (i.e., odds implied probability p_{odds} , and the games' final bet ratio, s_A). Odds have been found to be a good predictor of future events (Kuypers 2000). Note that s_A is not an unbiased estimation of event probability (Kain and Logan 2014). We use it simply for comparison.

BookMaker/Betting line: $p_A = p_{odds}$ (10)

BetRatio: $p_A = s_A$ (11)

The third set of baselines are reduced form models. Note that our proposed structural model eventually makes predictions based on odds and final bets. For comparison purposes, we experimented with the combination of p_{odds} and s_A using logit regression and derived four benchmark models (ReducedForm1 ~ ReducedForm4).

ReducedForm1: $logit(p_A) = \varphi_1 + \varphi_2 p_{odds}$ (12)

ReducedForm2: $logit(p_A) = \varphi_1 + \varphi_2 s_A$ (13)

ReducedForm3: $logit(p_A) = \varphi_1 + \varphi_2 p_{odds} + \varphi_3 s_A$ (14)

ReducedForm4: $logit(p_A) = \varphi_1 + \varphi_2 p_{odds} + \varphi_3 s_A + \varphi_4 p_{odds} s_A$ (15)

In prior research, Chen et al. (2011) built a complicated reduced form model customized for fixed odds betting by taking a simulation approach. We included this approach as a baseline, labeled as ReducedFormSimu.

Evaluation Metrics

We adopted widely accepted proper scoring rules (Winkler 1969) to evaluate future event forecasting performance, including the quadratic score rule (QSR), the logarithmic score rule (LSR), and the spherical score rule (SSR). When employing these metrics, we normalized these score rules so that the upper bound is always 100 (see Bickel 2007, 2010).

- Quadratic Score Rule (QSR) = $100 - 400 \times p_{lose}^2$, where p_{lose} is the probability assigned to the eventual losing team (that is, $p_{lose} = p_B$, if A happens, and $p_{lose} = p_A$, if B happens). The quadratic score is a linear transformation of the squared error and has been used in some betting websites. The upper bound of this metric is 100 and the lower bound is -300.
- Logarithmic Score Rule (LSR) = $100 + 144.27 \times \log(1 - p_{lose})$. The upper bound of this metric is 100 and the lower bound is $-\infty$.
- Spherical Score Rule (SSR) = $-241.42 + 341.42 \times (1 - p_{lose}) / \sqrt{p_A^2 + p_B^2}$. The upper bound of this metric is 100 and the lower bound is -241.42.

A prediction with a higher score on the three measures is more accurate.

Experimental Procedures

To compare the performance of our approach with that of InKlingmarkets.com, an auction-based prediction market, we used the events from the sina.com and sohu.com datasets to train the model and applied the learned model to generate predictions for the corresponding 26 events that were bet on at both InKlingmarkets and sina.com/sohu.com, as discussed earlier.

To compare our approach against the baseline methods, we took a bootstrapping approach. We conducted 500 rounds of experiments on each dataset. In each round, we randomly split the dataset into training and testing data. Finally, we conducted paired t-tests to compare the performance of these models.

Before the experimental runs, we first determined the polynomial order (D and Z) and estimated the crowd belief distribution function by estimating the parameters on the entire dataset. We report both the belief distributions specified by maximum likelihood and AIC criteria, named BD-ML and BD-AIC, respectively. After fixing D and Z , in each round of the experiments, we trained the coefficients of the belief distribution function using the training data and made predictions on the testing data. Following this procedure, the belief distributions varied across experiments in the prediction stage.

Results

Belief Distribution

By varying D and Z from 1 to 3, we obtained the optimal parameter $\theta \in \Theta$ for each setting of D and Z . Tables A1 through A3 in the Appendix report the log likelihood and AIC values for these models. The belief distribution functions in the models are also reported in the Appendix. As an example, on the Sina 2008 Olympic Games dataset, the maximum likelihood model has $D = 2$ and $Z = 2$ as follows:

$$f(x; p) = 0.74 \times \text{Beta}\left(x; 3.76 \times p + 0.1 \times p^2, 3.76 \times (1-p) + 0.1 \times (1-p)^2\right) + 0.26 \times \text{Beta}\left(x; 0.11 \times p + 66.32 \times p^2, 0.11 \times (1-p) + 66.32 \times (1-p)^2\right)$$

The AIC model combines three components with the smallest AIC ($D = 1, Z = 1$; $D = 1, Z = 2$; and $D = 1, Z = 3$) with the estimated belief distribution function:

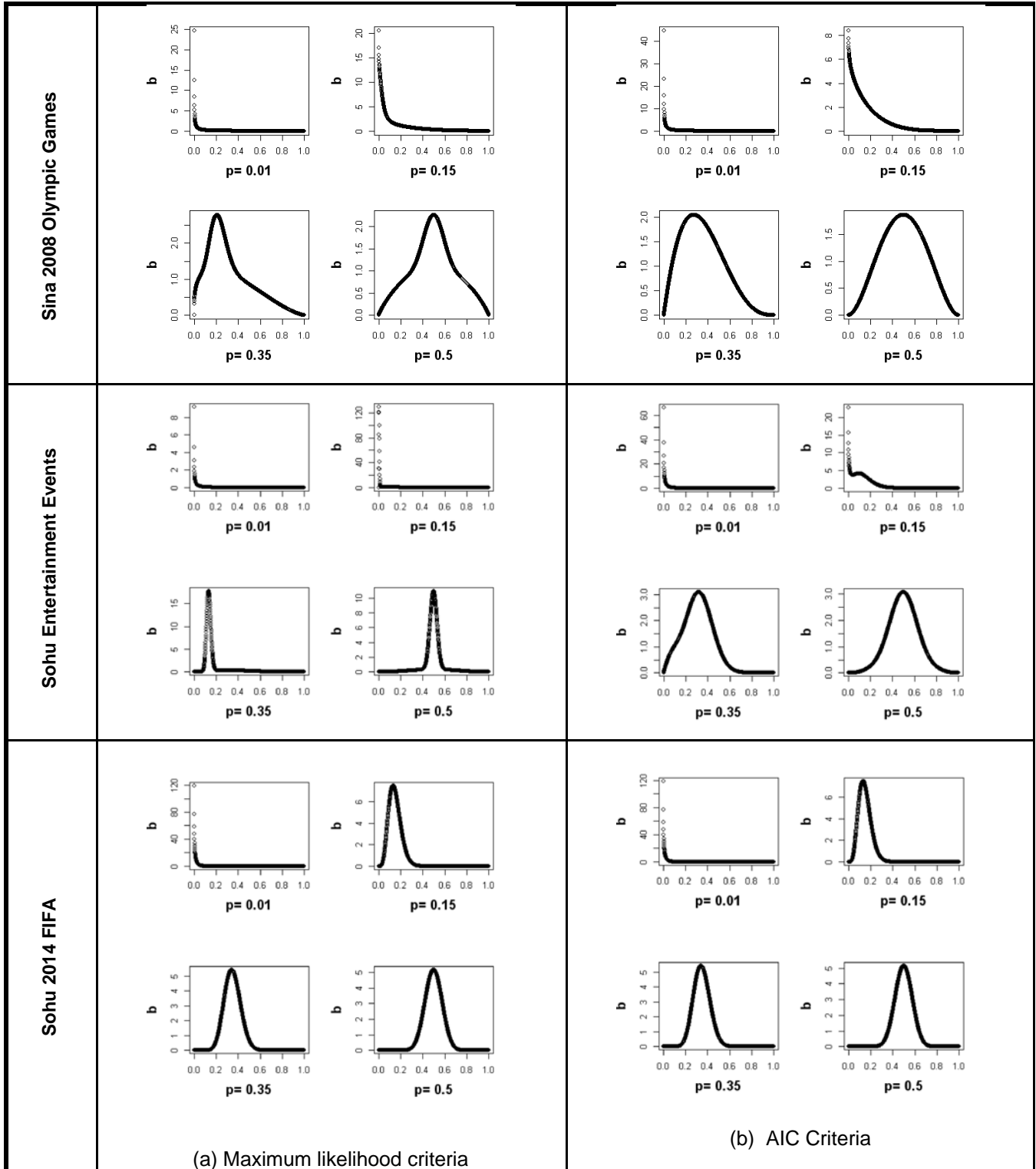


Figure 3. Belief Distributions under Different Event Probabilities

$$f(x; p) = 0.67 \times \text{Beta}(x; 5.98 \times p, 5.98 \times (1-p)) \\ + 0.24 \times \text{Beta}(x; 5.69 \times p + 0.16 \times p^2, 5.69 \times (1-p) + 0.16 \times (1-p)^2) \\ + 0.09 \times \text{Beta}(x; 5.95 \times p + 0.02 \times p^3, 5.95 \times (1-p) + 0.02 \times (1-p)^3)$$

We visualize the estimated belief distribution functions in Figure 3, in which event probability p varies from 0.01 to 0.5. If p is small, most participants do not believe that the event will happen, which follows a power law distribution. When p increases, the center of the crowd belief distribution tilts toward the right and the distribution shows characteristics of normal distribution. On the Sina 2008 Olympic Games and Sohu Entertainment datasets, the belief distributions estimated by the ML criteria have smaller variances than the belief distributions estimated by the AIC criteria. This shows that the models selected by the ML criteria are more sensitive to errors in the data or parameters. We also observe that the belief distribution associated with the ML criteria tends to peak closer to the edge than the event probability. One explanation is that some participants who held strong opinions affected the overall belief distribution. On the Sohu 2014 FIFA dataset, the two models resulted in very similar belief distributions and their means were very close to the event probabilities.

Predictive Power

Comparison with an Auction-Based Prediction Market

Table 1 shows the comparison between our proposed approach and InKlingmarkets.com. Our proposed methods, BD-ML and BD-AIC, have the highest SQR, LSR, and SSR in most cases. However, since the number of events for which direct comparisons are possible is very small, there is no significant statistical difference between these methods using pair-wise t-tests. Nonetheless, it is safe to claim that our proposed fixed odds betting-based prediction market achieves comparable performance with a representative auction-based prediction market.

Comparison with Other Baseline Approaches

Table 2 compares the predictive performance of our proposed models with bookmakers' estimations, final bet ratio, and the standard and customized reduced form models. The two BD models have the highest average scores among all models. Most of their scores are significantly better than the baseline models at the 99% confidence interval in paired t-tests.

Among the baseline methods, bookmakers' estimations, p_{odds} , generally perform well, showing the effectiveness of expert-based judgment and propitiatory models. This confirms what is known in the literature. Spann and Skiera (2009) found that auction-based prediction markets and betting odds perform equally well in making predictions. Note that, in our experiments, p_{odds} 's performance is less effective than that of the BD models. The experimental results also show final bet ratio is not a good predictor for future events, which is consistent with previous research (Kain and Logan 2014).

Among the reduced form models, those that combine multiple features (ReducedForm3 ~ ReducedForm4) generally outperform those using individual features (ReducedForm1 ~ ReducedForm2). The reduced form model customized for fixed odds betting predictions, ReducedFormSimu, generally has better performance than other reduced form models. However, the proposed BD models easily outperform ReducedFormSimu, illustrating the advantage of structural models over reduced form models. The proposed BD framework is a structural model that elaborates on participants' belief distributions.

Robustness Check

In the experiments reported above, for parameters operationalizing prospect theory, we specified τ as 0.69, ρ as 2.25, γ as 0.61, and q and p as 0.88, following the previous literature (Tversky and Kahneman 1992). To test the robustness of our approach against variations in these parameters, we varied them (± 0.1) and conducted additional experiments on the 243 parameter combinations with 50 rounds of bootstrapping each. Since it is difficult to visualize the results of all 243 settings in 5 dimensions, we plot performance changes for each parameter by computing averages over all other parameters, as shown in Figure 4.

Figure 4 shows the performance of BD-AIC and BD-ML under different parameter settings. For comparison, we also show the performance of the best benchmark in each dataset. Since the results on QSR, LSR, and SSR show similar patterns, we only show QSR in Figure 4. We observe that although the performance of the BD methods changed slightly with different prospect theory parameters, they consistently delivered much better performance than the benchmark algorithms. T-tests show that these differences are significant at the 90% confidence interval in all settings.

In terms of individual parameters' impact, the BD methods are relatively stable to parameters ρ , γ , and p . The changes in τ and q , which affect the shape of the weighting function and

Table 1. Comparison with an Auction-Based Prediction Market

Data	# Events	Model	Average # of Participants	QSR	LSR	SSR
Sina 2008 Olympic Games	6	Auction Market	59	-68.16	-70.94	-63.64
		BD-AIC	440	-71.06	-59.91	-73.71
		BD-ML	440	-67.57	-56.68	-70.33
Sohu Entertainment	2	Auction Market	160	-99.13	-164.67	-73.47
		BD-AIC	58	-43.98	-33.38	-49.62
		BD-ML	58	-9.14	-6.62	-10.98
Sohu 2014 FIFA	18	Auction Market	52	46.81	6.34	50.49
		BD-AIC	412	48.93	38.82	54.30
		BD-ML	412	48.91	38.81	54.28
Overall	26	Auction Market	62	9.05	-24.65	14.62
		BD-AIC	391	14.09	10.48	16.77
		BD-ML	391	17.56	13.28	20.50

Table 2. Comparison with Other Baseline Approaches

Model	Sina 2008 Olympic Games			Sohu Entertainment			Sohu 2014 FIFA		
	QSR	LSR	SSR	QSR	LSR	SSR	QSR	LSR	SSR
BookMaker/Betting line	9.10	6.74	10.62	0.10	-1.83	1.18	3.02	2.24	3.52
BetRatio	6.94	4.70	8.65	-32.75	-54.27	-26.11	-12.40	-33.65	-5.76
ReducedForm1	6.40	4.57	7.79	-5.07	-25.44	-2.49	3.08	2.88	3.02
ReducedForm2	7.87	5.86	9.15	-9.99	-8.55	-10.05	4.01	2.82	4.93
ReducedForm3	11.67	9.27	12.81	-10.98	-31.32	-8.37	5.13	4.15	5.71
ReducedForm4	10.91	8.90	11.91	-26.31	-100.05	-21.32	4.13	3.32	4.67
ReducedFormSimu	12.70	10.57	13.80	-3.21	3.02	-0.22	3.32	2.51	4.04
BD-AIC	16.10	12.99	17.51	5.36	4.57	5.66	5.81	4.34	6.74
BD-ML	16.33	13.19	17.76	11.35	9.99	12.02	5.81	4.34	6.74

Note: Values in bold are significantly higher than other values in the corresponding column at the 90% confidence interval.

valuation functions on gains, may lead to larger performance variations. On these two variables, the values recommended by (Tversky and Kahneman 1992) generally provided good but not top performance. On the Sohu Entertainment dataset, the BD-AIC and BD-ML methods had different responses to the value changes in these two parameters. On the Sina 2008 Olympic Games dataset, the BD-AIC algorithm achieved the best performance on our selected τ . In some application settings, such parameters may be specified based on human subject experiments.

Market Makers' Odds Setup and Fixed Odds Betting Prediction Performance

Our proposed BD framework taps into the predictive power of participants' opinions, a form of crowd wisdom. In our

model, we consider odds setup exogenous. Market makers' odds setups may be based on their proprietary models. Two interesting research questions arise in this context. In order to predict future events, how should market makers set the odds? Should they make an effort to improve their proprietary models and use more accurate predictions to set odds?

To fully address this problem would require new formal studies, beyond the scope of this paper. In this reported research, we aim to derive some preliminary computational insights using *post hoc* analysis. Figure 5 shows a scatter diagram of the BD-AIC model's performance (measured by QSR) to the correctness of p_{odds} in predicting future events (measured by p_{lose}) on the Sina 2008 Olympic Games dataset. There is a negative correlation between the two measures in the entire dataset (the solid line). The same kind of negative correlation can be observed if the dataset is separated ac-

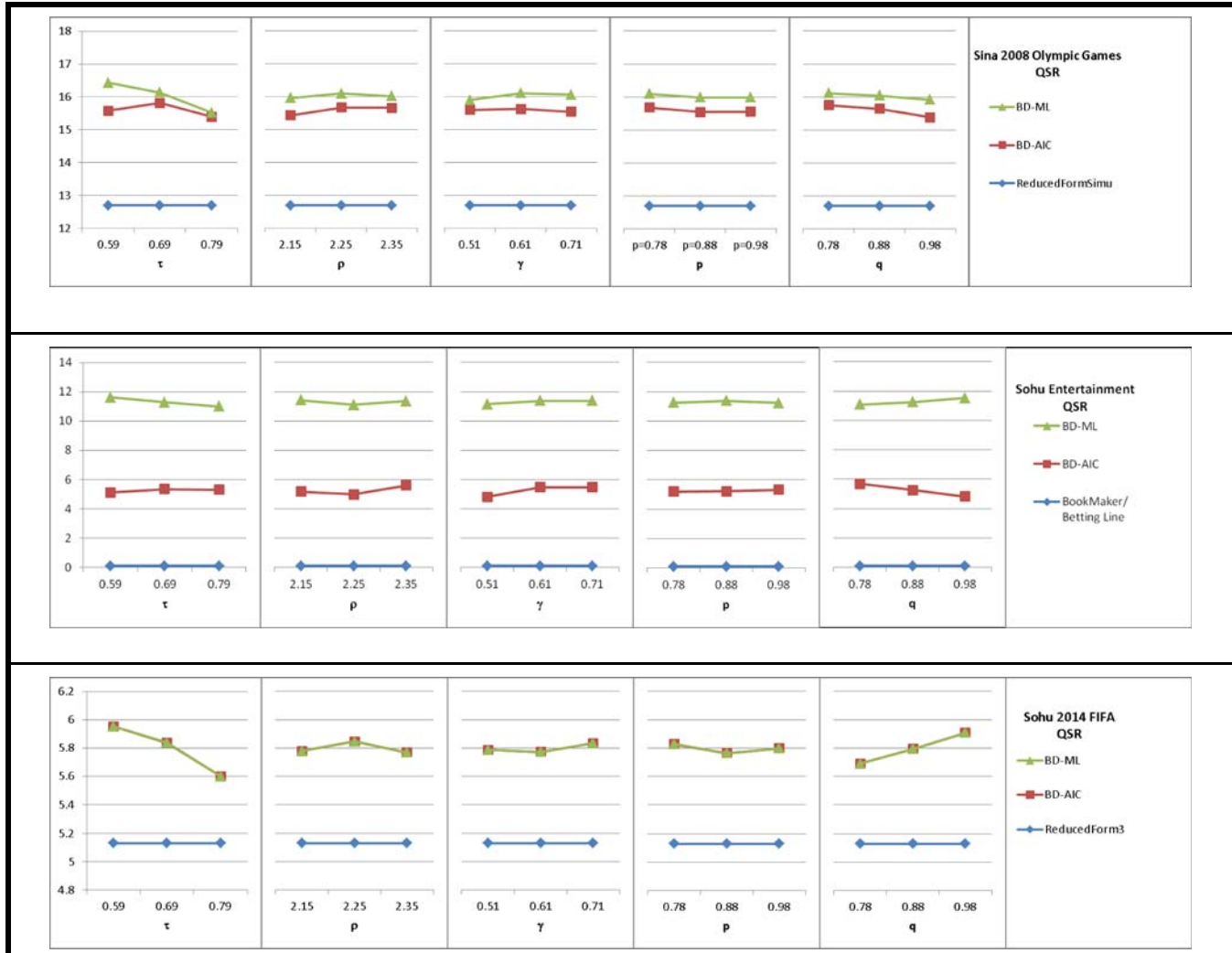


Figure 4. BD Models' Prediction Performance on Different Prospect Theory Parameters

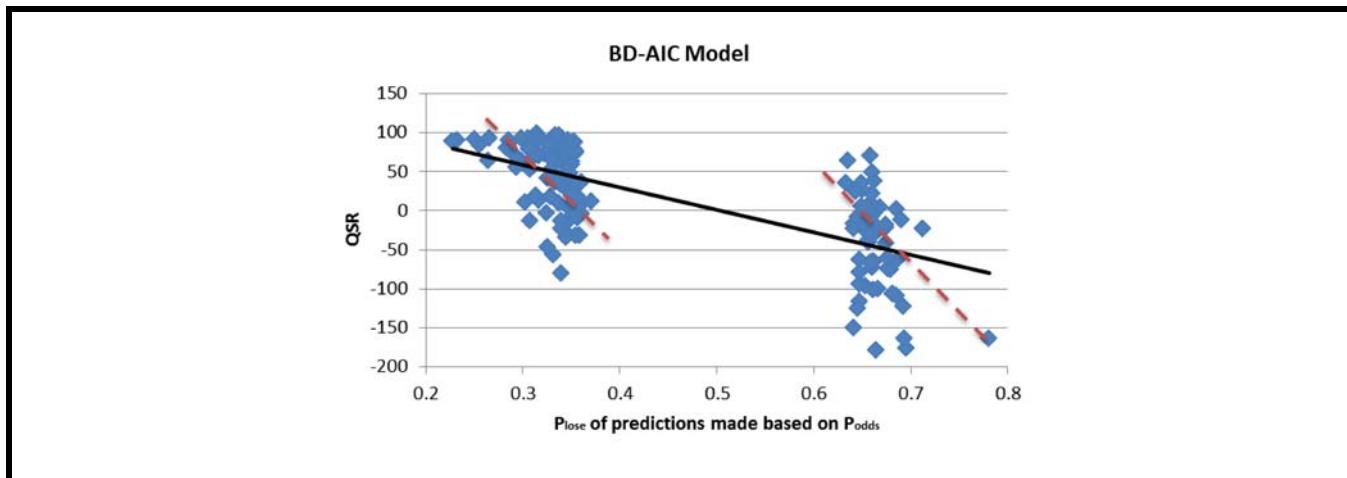


Figure 5. BD-AIC's Prediction Performance and the Correctness of p_{odds}

Table 3. Correlation Coefficients between BD Model's Prediction Performance and the Correctness of p_{odds} in Making Predictions

Data	Corr Coeff.	BD-ML			BD-AIC		
		QSR	LSR	SSR	QSR	LSR	SSR
Sina 2008 Olympic Games	All	-0.674	-0.662	-0.682	-0.704	-0.698	-0.708
	$p_{lose} < 0.5$	-0.707	-0.696	-0.716	-0.734	-0.728	-0.739
	$p_{lose} > 0.5$	-0.402	-0.401	-0.399	-0.411	-0.425	-0.396
Sohu Entertainment	All	-0.421	-0.377	-0.432	-0.476	-0.494	-0.452
	$p_{lose} < 0.5$	-0.246	-0.192	-0.265	-0.283	-0.278	-0.278
	$p_{lose} > 0.5$	-0.284	-0.287	-0.279	-0.284	-0.286	-0.279
Sohu 2014 FIFA	All	-0.507	-0.517	-0.497	-0.507	-0.516	-0.497
	$p_{lose} < 0.5$	-0.444	-0.457	-0.430	-0.443	-0.457	-0.429
	$p_{lose} > 0.5$	-0.284	-0.287	-0.279	-0.284	-0.286	-0.279

cording to $p_{lose} < 0.5$ (the two dot lines), i.e., separating correctly and incorrectly predicted events. This indicates that if the market maker sets a more accurate p_{odds} , the BD-AIC model's prediction would also be more accurate.

Table 3 extends this *post hoc* analysis to other performance measures for all of the available datasets in the form of correlation coefficients between p_{lose} (when using p_{odds} to predict future events) and BD models' performance. Since the events naturally break down to groups of correctly and incorrectly predicted events, the correlation coefficients are higher at the entire dataset level than within each group. Nevertheless, in all scenarios, we observe strong negative correlation between BD model performance and market makers' mistakes. There are two possible explanations. First, if the market makers have made an effort in estimating p_{odds} , such as in the Sina dataset, p_{lose} shows the difficulty of predicting the event. A larger p_{lose} tends to be associated with more difficult problems, with decreased performance of the BD models. Second, if the market makers have not made enough effort to estimate p_{odds} , such as in the two Sohu datasets, p_{lose} , to some extent, reflects the level of effort they made. A larger p_{lose} tends to appear on events with less accurate odds, which misleads the crowd and negatively impacts the performance of the BD models.

We also notice that if the models are built upon events with more participants, their prediction performance tends to be higher.

Discussion

The experimental findings indicate that our proposed BD framework significantly outperforms the reduced form models and achieves at least comparable performance with auction-

based prediction market mechanisms when predicting future events. Using fixed odds betting as a prediction market mechanism has significant implications for practitioners and researchers.

In Table 4, we compare the characteristics of traditional prediction markets with the fixed odds betting-based approach. In traditional prediction markets, the market price directly reflects the event probability. The fixed odds betting-based approach needs to develop an estimator to derive the event probability. The fixed odds betting-based prediction approach explicitly differentiates expert opinions (odds) from crowd opinions. The participants are forced to take these odds into consideration, which is modeled in the probability estimator. The crowds' belief can be estimated from crowd responses; if necessary, the decision maker can inspect experts' versus crowds' opinions to make decisions. A fixed odds betting-based prediction market requires less cognitive load from participants. Participants do not need to explicitly specify contract price. Instead, they just need to judge which side will lead to higher expected utility. We expect that this reduced cognitive load will likely result in a larger population of participants and more sources of information. To control the bookmaker's possible financial loss, participants can be rewarded based on the ranking of their number of correct predictions. The BD framework proposed in this paper is built on fixed odds and the betting ratio of the general public, which makes it more robust against market manipulation, especially price-based manipulation. The larger number of participants also makes manipulation more difficult.

Our proposed fixed odds betting-based prediction market provides an instrument to understand crowd beliefs under the influence of experts' opinions. In traditional prediction markets and the commonly used polling methods, expert opinions are not explicitly captured or modeled. Graefe (2015) found

Table 4. Comparison of Various Prediction Markets

	Fixed Odds	Traditional
Event probability	Calculated through an estimator	Price
Expert opinion	Explicit in odds	Together with the crowd
Cognitive load	Low	High
Reward rule	Ranking of correct predictions	# of correct predictions
Immune to manipulation	High	Low

that integrating information from prediction markets, expert judgments, and other methods provides more accurate predictions on German elections. Our study provides a direction to realize this kind of meaningful integration of multiple kinds of intelligence.

From an application perspective, the ability to differentiate experts' and crowds' opinions makes our approach particularly relevant to public administration and policy-making. In this study, our experiments are not based on public policy datasets due to data availability challenges. In practice, there have been several successful cases of applying prediction markets to predict political events, entertainment events, and sports events (Leigh and Wolfer 2007). Sonnemann et al. (2013) showed that people manifest similar psychological biases in prediction markets across application settings (e.g., on sports, economics, and horsing racing).

It is critical to note that prediction markets can only predict events that are external to the market (Abramowicz 2008). We cannot use market predictions to set a policy and then use the policy decision to reward participants. From this perspective, market participants need to follow a similar process (e.g., through social networks) in collecting information concerning the event (presumably outside of their control) for either sports or entertainment applications, or policy-related events. This shared information aggregation process across application contexts indicates the possibility of applying our approach to public policy applications. Of course, further studies are needed to establish the usefulness of our proposed approach in such policy-making contexts.

Our proposed approach has the following limitations. First, since our model is estimated from historical data, it requires the distributions of participants to be relatively stable across betting games. Second, the way that a fixed odds betting-based market is set up restricts experts from revising their opinions on an ongoing basis. Given the possibility of the presence of biases in odds setup, a fixed odds prediction market is more suitable for a short-span game. For long-running events, multiple fixed odds games could be constructed in a sequential manner. Third, in significant applica-

tions, when both traditional and fixed odds betting-based prediction markets are available for the same event, experienced participants may employ the information aggregated in the fixed odds betting-based prediction market to trade in a traditional prediction market. Such a cross-market information exchange may alter the effectiveness of our approach.

Conclusions

In this paper, we propose to repurpose fixed odds betting as an alternative prediction market mechanism, which has several desirable properties such as low cognitive overload imposed on participants and the ability to separate expert opinions and crowd wisdom. Since fixed odds betting lacks a mechanism to aggregate crowd intelligence, we developed a belief–decision (BD) framework along with a structural model as an event probability estimator. In our BD framework, crowd belief distribution is modeled by mixed Beta distributions and the participant's utility function is based on prospect theory. A maximum likelihood approach was developed to estimate the parameters of this BD framework from historical data and predict future event probabilities. We demonstrated the effectiveness of our approach using three real-world datasets. The results show that our approach outperforms bookmakers' estimations and reduced form models, and achieves comparable performance with auction-based prediction markets. In addition to event prediction, our proposed approach can also be used to estimate crowd belief distribution, which can be useful in many decision-making settings.

In our ongoing work, we are developing models that can be used to capture the dynamics of crowd activities. We are also working on extensions of our model to efficiently tackle more complicated decision scenarios than binary decisions.

As to further validation of our approach, our planned work is two-fold. First, we plan to conduct laboratory experiments to study participants' behavior (elicited through financial incentives) under various fixed odds betting scenarios. Second, we

have been investigating how fixed odds betting can be customized and embedded in selected business decision-making and policy-making scenarios. Such studies will help us gain better understanding of the applicability of fixed odds betting-based prediction market designs.

Acknowledgments

We would like to thank the senior editor, Kartik Hosanagar, as well as our anonymous associate editor and reviewers for their valuable suggestions that led to a considerable improvement of this paper. This work was supported by the National Natural Science Foundation of China (Grants 71025001, 71621002, 71572169), Guang Dong Natural Science Foundation (Grant 2015A030313876), Research Grants Council of the Hong Kong Special Administrative Region, China (Grant CityU 11503115), and the Peak Discipline Construction Project of Education at East China Normal University.

References

- Abramowicz, M. 2008. *Predictocracy: Market Mechanisms for Public and Private Decision Making*, New Haven, CT: Yale University Press.
- Abramowicz, M., and Henderson, T. 2007. "Prediction Markets for Corporate Governance," *Notre Dame Law Review* (82:4), pp. 1343-1414.
- Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., Levmore, S., Litan, R., Milgrom, P., Nelson, F. D., Neumann, G. R., Ottaviani, M., Schelling, T. C., Shiller, R. J., Smith, V. L., Snowberg, E., Sunstein, C. R., Tetlock, P. C., Tetlock, P. E., Varian, H. R., Wolfers, J., and Zitzewitz, E. 2008. "The Promise of Prediction Markets," *Science* (320:5878), pp. 877-878.
- Asch, P., Malkiel, B. G., and Quandt, R. E. 1982. "Racetrack Betting and Informed behavior," *Journal of Financial Economics* (10:2), pp. 187-194.
- Audun, J. 1997. "Artificial Reasoning with Subjective Logic," in *Proceedings of the 2nd Australian Workshop on Commonsense Reasoning*, Australian Computer Society, Perth, Australia.
- Berg, J. E., and Rietz, T. A. 2006. "The Iowa Electronic Markets: Stylized Facts and Open Issues," in *Information Markets: A New Way of Making Decisions*, R. W. Hahn and P. C. Tetlock (eds.), Washington, DC: AEI Press, pp. 142-169.
- Bickel, J. E. 2007. "Some Comparisons Among Quadratic, Spherical, and Logarithmic Scoring Rules," *Decision Analysis* (4:2), pp. 49-65.
- Bickel, J. E. 2010. "Scoring Rules and Decision Analysis Education," *Decision Analysis* (7:4), pp. 346-357.
- Blohm, I., Riedl, C., Leimeister, J. M., and Krcmar, H. 2011. "Idea Evaluation Mechanisms for Collective Intelligence in Open Innovation Communities: Do Traders Outperform Raters?," in *Proceedings of the 32nd International Conference on Information Systems*, Shanghai, China.
- Blume, M., Luckner, S., and Weinhardt, C. 2010. "Fraud Detection in Play-Money Prediction Markets," *Information Systems and E-Business Management* (8:4), pp. 395-413.
- Bondarenko, O., and Bossaerts, P. 2000. "Expectations and Learning in Iowa," *Journal of Banking and Finance* (24), pp. 1535-1555.
- Burnham, K. P., and Anderson, D. R. 2002. *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach* (2nd ed.), New York: Springer.
- Casella, G., and Berger, R. 2002. *Statistical Inference* (2nd ed.), Boston: Brooks/Cole Publishing.
- Chen, W., Li, X., and Zeng, D. 2011. "Estimating Collective Belief in Fixed Odds Betting," in *Proceedings of the Pacific Asia Workshop on Intelligence and Security Informatics*, Beijing, pp. 54-63.
- Chen, Y., Dimitrov, S., Sami, R., Reeves, D. M., Pennock, D. M., Hanson, R. D., Fortnow, L., and Gonen, R. 2010. "Gaming Prediction Markets: Equilibrium Strategies with a Market Maker," *Algorithmica* (58:4), pp. 930-969.
- Chintagunta, P., Erdem, T., Rossi, P. E., and Wedel, M. 2006. "Structural Modeling in Marketing: Review and Assessment," *Marketing Science* (25:6), pp. 604-616.
- Dahan, E., Soukhoroukova, A., and Spann, M. 2010. "New Product Development 2.0: Preference Markets—How Scalable Securities Markets Identify Winning Product Concepts and Attributes," *Journal of Product Innovation Management* (27:7), pp. 937-954.
- Dai, W., Jin, G., Lee, J., and Luca, M. 2012. "Optimal Aggregation of Consumer Ratings: An Application to Yelp.com," unpublished paper, University of Maryland.
- Erdem, T., and Keane, M. P. 1996. "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science* (15:1), pp. 1-20.
- Erikson, R. S., and Wlezien, C. 2008. "Are Political Markets Really Superior to Polls as Election Predictors?," *Public Opinion Quarterly* (72:2), pp. 190-215.
- Forrest, D., Goddard, J., and Simmons, R. 2005. "Odds-Setters as Forecasters: The Case of English Football," *International Journal of Forecasting* (21:3), pp. 551-564.
- Forsythe, R., Nelson, F., Neumann, G. R., and Wright, J. 1992. "Anatomy of an Experimental Political Stock Market," *American Economic Review* (82:5), pp. 1142-1161.
- Gabriel, P. E., and Marsden, J. R. 1990. "An Examination of Market-Efficiency in British Racetrack Betting," *Journal of Political Economy* (98:4), pp. 874-885.
- Goddard, J., and Asimakopoulos, I. 2004. "Forecasting Football Results and the Efficiency of Fixed-Odds Betting," *Journal of Forecasting* (23:1), pp. 51-66.
- Graefe, A. 2015. "German Election Forecasting: Comparing and Combining Methods for 2013," *German Politics* (24:2), pp. 195-204.
- Graefe, A., and Armstrong, J. S. 2011. "Comparing Face-to-Face Meetings, Nominal Groups, Delphi and Prediction Markets on an Estimation Task," *International Journal of Forecasting* (27:1), pp. 183-195.

- Green, K. C., Armstrong, J. S., and Graefe, A. 2007. "Methods to Elicit Forecasts from Groups: Delphi and Prediction Markets Compared," *Foresight: The International Journal of Applied Forecasting* (8), pp. 17-20.
- Gruca, T. S., Berg, J., and Cipriano, M. 2003. "The Effect of Electronic Markets on Forecasts of New Product Success," *Information Systems Frontiers* (5:1), pp. 95-105.
- Guo, Z. L., Fang, F., and Whinston, A. B. 2006. "Supply Chain Information Sharing in a Macro Prediction Market," *Decision Support Systems* (42:3), pp. 1944-1958.
- Hansen, J., Schmidt, C., and Strobel, M. 2004. "Manipulation in Political Stock Markets—Preconditions and Evidence," *Applied Economics Letters* (11:7), pp. 459-463.
- Hanson, R. D. 2003. "Combinatorial Information Market Design," *Information Systems Frontiers* (5:1), pp. 107-119.
- Hanson, R. D. 2006a. "Decision Markets for Policy Advice," in *Promoting the General Welfare: New Perspectives on Government Performance*, E. Patashnik and A. Gerber (eds.), Washington, DC: Brookings Institution Press, pp. 151-173.
- Hanson, R. D. 2006b. "Designing Real Terrorism Futures," *Public Choice* (128:1), pp. 257-274.
- Healy, P. J., Linardi, S., Lowery, J. R., and Ledyard, J. O. 2010. "Prediction Markets: Alternative Mechanisms for Complex Environments with Few Traders," *Management Science* (56:11), pp. 1977-1996.
- Jian, L., and Sami, R. 2012. "Aggregation and Manipulation in Prediction Markets: Effects of Trading Mechanism and Information Distribution," *Management Science* (58:1), pp. 123-140.
- Jullien, B., and Salanie, B. 2000. "Estimating Preferences Under Risk: The Case of Racetrack Bettors," *Journal of Political Economy* (108:3), pp. 503-530.
- Kain, K. J., and Logan, T. D. 2014. "Are Sports Betting Markets Prediction Markets? Evidence From a New Test," *Journal of Sports Economics* (15:1), pp. 45-63.
- Kelley, C. T. 1987. *Iterative Methods for Optimization*, Philadelphia: Society for Industrial Mathematics.
- Kim, M. 1998. "Convergence of the Nelder–Mead Simplex Method to a Nonstationary Point," *SIAM Journal on Optimization* (9:1), pp. 148-158.
- Kuypers, T. 2000. "Information and Efficiency: An Empirical Study of a Fixed Odds Betting Market," *Applied Economics* (32:11), pp. 1353-1363.
- Lagarias, J., Reeds, J., Wright, M., and Wright, P. 1998. "Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions," *SIAM Journal on Optimization* (9:1), pp. 112-147.
- Lee, D. S., and Moretti, E. 2009. "Bayesian Learning and the Pricing of New Information: Evidence from Prediction Markets," *The American Economic Review* (99:2), pp. 330-336.
- Leigh, A., and Wolfer, J. 2007. "Prediction Markets for Business and Public Policy," *The Melbourne Review* (3:1), pp. 7-15.
- Madhavan, A. 2012. "Trading Mechanisms in Securities Markets," *Journal of Finance* (47:2), pp. 607-641.
- Meirowitz, A., and Tucker, J. A. 2004. "Learning from Terrorism Markets," *Perspectives on Politics* (2:2), pp. 331-335.
- Onkal, D., Goodwin, P., Thomson, M., Gonul, S., and Pollock, A. 2009. "The Relative Influence of Advice from Human Experts and Statistical Methods on Forecast Adjustments," *Journal of Behavioral Decision Making* (22:4), pp. 390-409.
- Parkes, D., Ungar, L., and Foster, D. 1999. "Accounting for Cognitive Costs in On-Line Auction Design," in *Agent Mediated Electronic Commerce: First International Workshop on Agent-Mediated Electronic Trading*, P. Noriega and C. Sierra (eds.), New York: Springer, pp. 25-40.
- Pennock, D. M. 2004. "A Dynamic Pari-Mutuel Market for Hedging, Wagering, and Information Aggregation," in *Proceedings of the ACM Conference on Electronic Commerce*, New York: ACM Press, pp. 170-179.
- Polgreen, P. M., Nelson, F. D., Neumann, G. R., and Weinstein, R. A. 2007. "Use of Prediction Markets to Forecast Infectious Disease Activity," *Clinical Infectious Diseases* (44:2), pp. 272-279.
- Quandt, R. E. 1986. "Betting and Equilibrium," *The Quarterly Journal of Economics* (101:1), pp. 201-207.
- Renn, O., Webler, T., Rakel, H., Diemel, P., and Johnson, B. 1993. "Public Participation in Decision Making: A Three-Step Procedure," *Policy Sciences* (26:3), pp. 189-214.
- Rhode, P. W., and Strumpf, K. S. 2008. "Manipulating Political Stock Markets: A Field Experiment and a Century of Observational Data," unpublished paper, University of North Carolina at Chapel Hill.
- Ritterman, J., Osborne, M., and Klein, E. 2009. "Using Prediction Markets and Twitter to Predict a Swine Flu Pandemic," in *Proceedings of the 1st International Workshop on Mining Social Media*, Sevilla, Spain, pp. 1-9.
- Rothschild, D., and Wolfers, J. 2008. "Market Manipulation Muddies Election Outlook," *The Wall Street Journal*, October 2.
- Sauer, R. D. 1998. "The Economics of Wagering Markets," *Journal of Economic Literature* (36:4), pp. 2021-2064.
- Schreiber, E. S. 2009. "Prediction Markets: Trading Uncertainty for Collective Wisdom," in *Proceedings of the International Studies Association Annual Conference*, New York, pp. 1-16.
- Servan-Schreiber, E., Wolfers, J., Pennock, D. M., and Galebach, B. 2004. "Prediction Markets: Does Money Matter?," *Electronic Markets* (14:3), pp. 243-251.
- Song, C., Boulier, B. L., and Stekler, H. O. 2007. "The Comparative Accuracy of Judgmental and Model Forecasts of American Football Games," *International Journal of Forecasting* (23:3), pp. 405-413.
- Sonnemann, U., Camerer, C. F., Fox, C. R., and Langer, T. 2013. "How Psychological Framing Affects Economic Market Prices in the Lab and Field," *Proceedings of the National Academy of Sciences of the United States of America* (110:29), pp. 11779-11784.
- Spann, M., and Skiera, B. 2009. "Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters," *Journal of Forecasting* (28:1), pp. 55-72.
- Steven, D. L. 2004. "Why Are Gambling Markets Organised So Differently from Financial Markets?," *Economic Journal* (114:495), pp. 223-246.
- Strumbelj, E., and Sikonja, M. R. 2010. "Online Bookmakers' Odds as Forecasts: The Case of European Soccer Leagues," *International Journal of Forecasting* (26:3), pp. 482-488.

- Tversky, A., and Kahneman, D. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* (5:4), pp. 297-323.
- Winkler, R. L. 1969. "Scoring Rules and the Evaluation of Probability Assessors," *Journal of the American Statistical Association* (64:327), pp. 1073-1078.
- Wolfers, J., and Zitzewitz, E. 2004. "Prediction Markets," *Journal of Economic Perspectives* (18:2), pp. 107-126.
- Wolfers, J., and Zitzewitz, E. 2006. "Five Open Questions About Prediction Markets," in *Information Markets: A New Way of Making Decisions*, R. W. Hahn and P. C. Tetlock (eds.), Washington, DC: AEI Press, pp. 13-36.
- Woodland, L. M., and Woodland, B. M. 1994. "Market-Efficiency and the Favorite-Longshot Bias—The Baseball Betting Market," *Journal of Finance* (49:1), pp. 269-279.
- Zhao, Y., Yang, S., Narayan, V., and Zhao, Y. 2013. "Modeling Consumer Learning from Online Product Reviews," *Marketing Science* (32:1), pp. 153-169.

About the Authors

Weiyun Chen is an assistant professor in the Department of Educational Information Technology at East China Normal University. His current research focuses on educational data mining and social computing. He received his Ph.D. in Control Theory and Control Engineering from Chinese Academy of Sciences, and his Bachelor's and Master's degrees from the Department of Automation at Huazhong University of Science and Technology, China. His research has appeared in *IEEE Intelligent Systems*, *Overseas Scholars*, and *Progress in Research on Information Management and Information System* (book chapter, in Chinese). He has also published several research articles in refereed conference proceedings including the Workshop on Information Technology and Systems, International Workshop on Social Computing, the Pacific and Asian Workshop on Intelligence and Security Informa-

tics, Chinese Conference on Social Computing and American Educational Research Association Annual Meeting.

Xin Li is an associate professor in the Department of Information Systems at the City University of Hong Kong. He received his Ph.D. in Management Information Systems from the University of Arizona, and his Bachelor's and Master's degrees from the Department of Automation at Tsinghua University, China. His research interests include business intelligence and knowledge discovery, social network analysis, social media, and e-commerce. His work has appeared in *INFORMS Journal on Computing*, *Journal of Management Information Systems*, *Decision Support Systems*, *ACM Transactions on Management Information Systems*, *Journal of the American Society for Information Science and Technology*, and various IEEE Transactions, among other venues.

Daniel Zeng is Gentile Family Professor in the Department of Management Information Systems at the University of Arizona. He is an affiliated research fellow at the Institute of Automation, Chinese Academy of Sciences. His current research interests include intelligence and security informatics, infectious disease informatics, social computing, recommender systems, software agents, spatial-temporal data analysis, business analytics, and online advertising. He received M.S. and Ph.D. degrees in industrial administration from Carnegie Mellon University and a B.S. degree in economics and operations research from the University of Science and Technology of China, Hefei, China. He has published one monograph and more than 300 peer-reviewed articles. He serves as the editor in chief of *IEEE Intelligent Systems* and is a fellow of IEEE. His research has been funded mainly by the U.S. National Science Foundation, the U.S. National Institutes of Health, the U.S. Department of Homeland Security, the National Natural Science Foundation of China, and the Ministry of Health of China. As principal investigator, he has received more than \$20 million in government research support. He is president of the IEEE Intelligent Transportation Systems Society and the past chair of INFORMS College on Artificial Intelligence.

MODELING FIXED ODDS BETTING FOR FUTURE EVENT PREDICTION

Weiyun Chen

Department of Educational Information Technology, Faculty of Education, East China Normal University, Shanghai, CHINA {weiyun.chen@qq.com}

Xin Li

Department of Information Systems, College of Business, City University of Hong Kong, Hong Kong, CHINA {xin.li.phd@gmail.com}

Daniel Zeng

Department of Management Information Systems, University of Arizona, Tucson, AZ 85721 U.S.A., and State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, CHINA {zeng@email.arizona.edu}

Appendix

Theorem 1: If $\alpha_i(p) = \beta_i(1-p)$, then $f(x; p, \theta) = f(1-x; 1-p, \theta)$.

Proof:

$$\begin{aligned} f(x; p, \theta) &= \sum_{i=1}^D \frac{\lambda_i}{\sum_{j=1}^D \lambda_j} \cdot \text{Beta}(x; \alpha_i(p), \beta_i(p)) \\ &= \sum_{i=1}^D \frac{\lambda_i}{\sum_{j=1}^D \lambda_j} \cdot \text{Beta}(1-x; \beta_i(p), \alpha_i(p)) \\ &= \sum_{i=1}^D \frac{\lambda_i}{\sum_{j=1}^D \lambda_j} \cdot \text{Beta}(1-x; \alpha_i(1-p), \beta_i(1-p)) = f(1-x; 1-p, \theta) \end{aligned}$$

Theorem 2: $E[\text{Beta}(x; \alpha_i(p), \beta_i(p))]$ is strictly increasing with p .

Proof:

$$E[\text{Beta}(x; \alpha_i(p), \beta_i(p))] = \frac{\alpha_i(p)}{\alpha_i(p) + \beta_i(p)} = \frac{1}{1 + \frac{\beta_i(p)}{\alpha_i(p)}}$$

Since p is within $(0, 1)$, both $\alpha_i(p)$ and $\beta_i(p)$ are positive.

$$\frac{\partial \frac{\beta_i(p)}{\alpha_i(p)}}{\partial p} = \frac{\alpha_i(p) \frac{\partial \beta_i(p)}{\partial p} - \beta_i(p) \frac{\partial \alpha_i(p)}{\partial p}}{\alpha_i(p)^2} = \frac{\alpha_i(p) \left[-u_{i1} - \sum_{h=2}^Z u_{ih} (1 - p_A)^{h-1} \right] - \beta_i(p) \left[u_{i1} + \sum_{h=2}^Z u_{ih} p_A^{h-1} \right]}{\alpha_i(p)^2}$$

Clearly, $\frac{\partial \frac{\beta_i(p)}{\alpha_i(p)}}{\partial p} \leq 0$ and the equal sign holds only if all $u_{ih}=0$ when p is within range $(0, 1)$.

If all $u_{ih}=0$, both $\alpha_i(p)$ and $\beta_i(p)$ equal 0, which is in conflict with our assumptions.

Thus $\frac{\beta_i(p)}{\alpha_i(p)}$ is strictly decreasing in p , and $\frac{\alpha_i(p)}{\alpha_i(p) + \beta_i(p)}$ is strictly increasing with p . ■

Theorem 3: If parameters $q, p, \rho, \gamma, \tau, o_A, o_B$ are all positive, there exists and only exists one belief value $c \in (0,1)$, called balance belief hereafter, satisfying $U(c, o_A) = U(1 - c, o_B)$.

Proof: When q, p, ρ, γ, τ are positive, both $w^+(\cdot)$ and $w^-(\cdot)$ are strictly increasing functions. Accordingly, the utility function $U(x, o)$ is a strictly increasing function and $U(1 - x, o)$ is a strictly decreasing function in belief x . Given $o_A > 0$ and $o_B > 0$, $[U(x, o_A) - U(1-x, o_B)]$ is strictly increasing. It is easy to verify that $U(x = 0, o_A) < 0 < U(x = 1, o_B)$ and $U(x = 1, o_A) > 0 > U(x = 0, o_B)$. Thus, $[U(x, o_A) - U(1-x, o_B)] < 0$ for $x = 0$ and $[U(x, o_A) - U(1 - x, o_B)] > 0$ for $x = 1$. As such, there must exist one and only one balance belief $x = c$, satisfying $U(c, o_A) = U(1 - c, o_B)$. ■

Theorem 4: For sufficiently large m_i , maximizing equation 7 reduces to solving $PA(p_i, \theta) = s_{iA}$.

Proof:

$Lc(p, \theta)$ in equation (7) is continuous and differentiable. Since $0 < p_i < 1$, the value of p_i that maximizes $Lc(p, \theta)$, if it exists, must satisfy the first-order condition $\frac{\partial Lc(p_i, \theta)}{\partial p_i} = 0$.

$$\begin{aligned} & \frac{\partial Lc(p_i, \theta)}{\partial p_i} \\ &= m_i s_{iA} \frac{1}{PA(p_i, \theta)} \frac{\partial PA(p_i, \theta)}{\partial p_i} + m_i (1 - s_{iA}) \frac{1}{1 - PA(p_i, \theta)} \frac{-\partial PA(p_i, \theta)}{\partial p_i} + R_i \frac{1}{p_i} + (1 - R_i) \frac{-1}{1 - p_i} \\ &= m_i \frac{\partial PA(p_i, \theta)}{\partial p_i} \left(\frac{s_{iA}}{PA(p_i, \theta)} - \frac{(1 - s_{iA})}{1 - PA(p_i, \theta)} \right) + \left[\frac{R_i}{p_i} - \frac{(1 - R_i)}{1 - p_i} \right] \\ &= m_i \frac{\partial PA(p_i, \theta)}{\partial p_i} \left(\frac{s_{iA} (1 - PA(p_i, \theta)) - (1 - s_{iA}) PA(p_i, \theta)}{PA(p_i, \theta) (1 - PA(p_i, \theta))} \right) + \left[\frac{R_i (1 - p_i) - (1 - R_i) p_i}{p_i (1 - p_i)} \right] \\ &= m_i \frac{\partial PA(p_i, \theta)}{\partial p_i} \left(\frac{s_{iA} - PA(p_i, \theta)}{PA(p_i, \theta) (1 - PA(p_i, \theta))} \right) + \frac{R_i - p_i}{p_i (1 - p_i)} = 0 \end{aligned}$$

Namely: $\left[PA(p_i, \theta) - s_{iA} \right] \frac{\partial PA(p_i, \theta)}{\partial p_i} = \frac{PA(p_i, \theta) (1 - PA(p_i, \theta)) (R_i - p_i)}{m_i p_i (1 - p_i)}$

$$\lim_{m_i \rightarrow +\infty} [PA(p_i, \theta) - s_{iA}] \frac{\partial PA(p_i, \theta)}{\partial p_i} = \lim_{m_i \rightarrow +\infty} \left[\frac{PA(p_i, \theta)(1 - PA(p_i, \theta))(R_i - p_i)}{m_i p_i (1 - p_i)} \right] = 0$$

According to Theorem 4, $\frac{\partial PA(p_i, \theta)}{\partial p_i} > 0$, we obtain $\lim_{m_i \rightarrow +\infty} [PA(p_i, \theta) - s_{iA}] = 0$. ■

Theorem 5: $I_c(\alpha(p), \beta(p))$ is strictly decreasing in p , where $I_c(\alpha(p), \beta(p))$ is the regularized incomplete Beta function $\int_0^c Beta(x; \alpha(p), \beta(p)) dx$.

Proof: Based on the chain rule of multivariable calculus, $\frac{\partial I_c(\alpha(p), \beta(p))}{\partial p} = \frac{\partial I_c(\alpha(p), \beta(p))}{\partial \alpha} \frac{\partial \alpha}{\partial p} + \frac{\partial I_c(\alpha(p), \beta(p))}{\partial \beta} \frac{\partial \beta}{\partial p}$.

$$\frac{\partial I_c(\alpha, \beta)}{\partial \alpha} = [\log(c) - \varphi(\alpha) + \varphi(\alpha + \beta)] I_c(\alpha, \beta) - \frac{\Gamma(\alpha)\Gamma(\alpha + \beta)}{\Gamma(\beta)} c^\alpha \sum_{k=0}^{\infty} \frac{(\alpha)_k (\alpha)_k (1 - \beta)_k c^k}{k! \Gamma(k + 1 + \alpha) \Gamma(k + 1 + \beta)}$$

where $(\cdot)_k$ is the Pochhammer symbol specified as $(x)_0 = 1$; $(x)_n = x(x + 1)(x + 2) \dots (x + n - 1)$.

Since $\varphi(\alpha) = \sum_{k=1}^{\infty} \left(\frac{1}{k} - \frac{1}{k + \alpha - 1} \right) - r$, $\varphi(\alpha + \beta) - \varphi(\alpha) = \sum_{k=1}^{\infty} \left(\frac{1}{k + \alpha + \beta - 1} - \frac{1}{k + \alpha - 1} \right) < 0$ when $\alpha > 0$ and $\beta > 0$.

Since $c < 1$, $\log(c) < 0$ and $[\log(c) - \varphi(\alpha) + \varphi(\alpha + \beta)] < 0$.

Since $I_c(\alpha, \beta) > 0$ if $0 < c < 1$, we have $[\log(c) - \varphi(\alpha) + \varphi(\alpha + \beta)] I_c(\alpha, \beta) < 0$.

Since $\Gamma(x) > 0$ if $x > 0$, we have $\frac{\Gamma(\alpha)\Gamma(\alpha + \beta)}{\Gamma(\beta)} c^\alpha \sum_{k=0}^{\infty} \frac{(\alpha)_k (\alpha)_k (1 - \beta)_k c^k}{k! \Gamma(k + 1 + \alpha) \Gamma(k + 1 + \beta)} > 0$

Thus, $\frac{\partial I_c(\alpha, \beta)}{\partial \alpha} < 0$ when $0 < c < 1$.

Similarly, we can prove $\frac{\partial I_c(\alpha, \beta)}{\partial \beta} > 0$ when $0 < c < 1$.

It is clear $\frac{\partial \alpha}{\partial p} > 0$ and $\frac{\partial \beta}{\partial p} < 0$ when $0 < p < 1$.

Thus, $\frac{\partial I_c(\alpha(p), \beta(p))}{\partial p} < 0$, i.e., $I_c(\alpha(p), \beta(p))$ is strictly decreasing in p . ■

Formula for AIC

The value of AIC criteria is computed as

$$AICc = 2t + 2t(t + 1) / (N - t - 1) - 2Lc(\theta)$$

where N denotes the number of data instances and t denotes the number of parameters, which is

$$t = \begin{cases} D(1 + Z) & D > 1 \\ Z & D = 1 \end{cases}$$

according to equations (1) and (2).

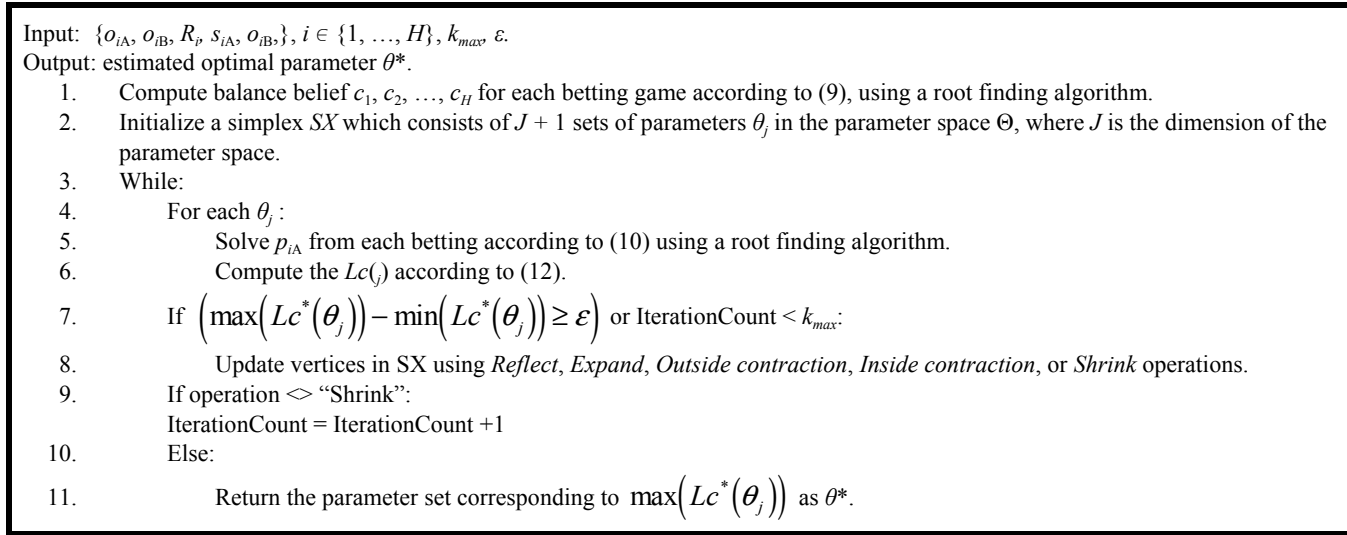


Figure A1. Maximum Likelihood Estimation Using a Nelder–Mead Method

Detailed Belief Distribution Estimation Procedure

Sina 2008 Olympic Games Dataset

For each setting of D and Z , varying from 1 to 3, respectively, we numerically obtained the optimal parameters $\theta^* \in \Theta$. Table A1 reports the log likelihood and AIC values for these models. Generally, the model’s likelihood converges when D and Z are larger than 2. The model with $D = 2$ and $Z = 2$ is the model with the maximum likelihood. The estimated belief distribution function is given as:

$$f(x; p) = 0.74 * \text{Beta}(x; 3.76 * p + 0.1 * p^2, 3.76 * (1 - p) + 0.1 * (1 - p)^2) + 0.26 * \text{Beta}(x; 0.11 * p + 66.32 * p^2, 0.11 * (1 - p) + 66.32 * (1 - p)^2)$$

For the AIC criteria, we combined the three components with smallest $AICc$ ($D = 1, Z = 1; D = 1, Z = 2; \text{ and } D = 1, Z = 3$). The estimated belief distribution function is given as

$$f(x; p) = 0.67 * \text{Beta}(x; 5.98 * p, 5.98 * (1 - p)) + 0.24 * \text{Beta}(x; 5.69 * p + 0.16 * p^2, 5.69 * (1 - p) + 0.16 * (1 - p)^2) + 0.09 * \text{Beta}(x; 5.95 * p + 0.02 * p^3, 5.95 * (1 - p) + 0.02 * (1 - p)^3)$$

Table A1. Log Likelihood and AIC Values (Sina 2008 Olympic Games)						
D	Z = 1		Z = 2		Z = 3	
	Log likelihood	AIC	Log likelihood	AIC	Log likelihood	AIC
1	-101.12	204.26	-101.11	206.29	-101.12	208.38
2	-101.03	210.30	-100.79	214.11	-101.05	219.01
3	-101.03	214.59	-100.80	220.74	-100.82	227.67

Sohu Entertainment Dataset

Table A2 shows the results on the Sohu entertainment event dataset, varying D and Z from 1 to 3. The model ($D = 3$ and $Z = 3$) is the model with the maximum likelihood. The estimated belief distribution function is as follows:

$$f(x; p) = 0.84 * Beta(x; 0.1 * p + 0.1 * p^2 + 1000 * p^3, 0.1 * (1 - p) + 0.1 * (1 - p)^2 + 1000 * (1 - p)^3) + 0.16 * Beta(x; 7.54 * p + 0.1 * p^2 + 0.1 * p^3, 7.54 * (1 - p) + 0.1 * (1 - p)^2 + 0.1 * (1 - p)^3)$$

For the AIC criteria, we combined the three components with smallest AIC ($D = 1, Z = 1$ and $D = 1, Z = 2$ and $D = 1, Z = 3$). The estimated belief distribution function is given as

$$f(x; p) = 0.62 * Beta(x; 21.28 * p, 21.28 * (1 - p)) + 0.29 * Beta(x; 1.29 * p + 13.3 * p^2, 1.29 * (1 - p) + 13.3 * (1 - p)^2) + 0.09 * Beta(x; p + 12.6 * p^2 + p^3, (1 - p) + 12.6 * (1 - p)^2 + (1 - p)^3)$$

Table A2. Likelihood and AIC Values (Sohu Entertainment)						
D	$Z = 1$		$Z = 2$		$Z = 3$	
	Log likelihood	AIC	Log likelihood	AIC	Log likelihood	AIC
1	-22.56	47.25	-22.20	48.79	-22.20	51.20
2	-22.26	53.90	-21.32	57.75	-21.32	64.40
3	-22.26	59.63	-21.32	68.14	-21.20	81.26

Sohu 2014 FIFA Dataset

Table A3 shows the results on the Sohu 2014 FIFA dataset, varying D and Z from 1 to 3. The model ($D = 1$ and $Z = 1$) is the model with the maximum likelihood. The estimated belief distribution is given as

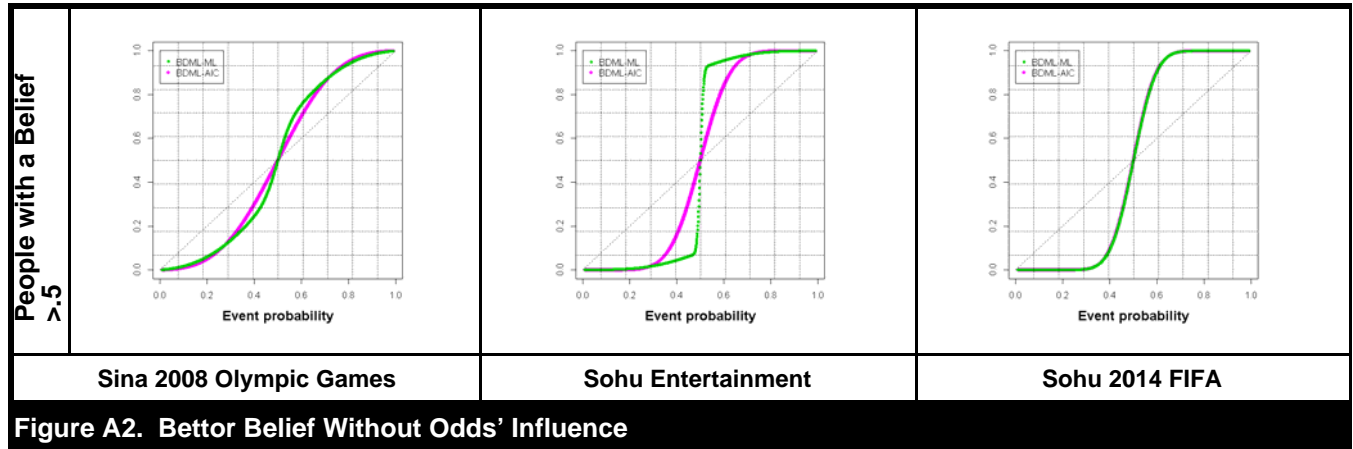
$$f(x; p) = Beta(x; 42.18 * p, 42.18 * (1 - p))$$

For the AIC criteria, we combine the three components with smallest AIC ($D = 1, Z = 1$; $D = 1, Z = 2$; and $D = 1, Z = 3$). The estimated belief distribution function is given as

$$f(x; p) = 0.67 * Beta(x; 42.18 * p, 42.18 * (1 - p)) + 0.24 * Beta(x; 41.88 * p, 41.88 * (1 - p)) + 0.09 * Beta(x; 41.98 * p, 41.98 * (1 - p))$$

Table A-3: Likelihood and AIC Values (Sohu 2014 FIFA)						
D	$Z = 1$		$Z = 2$		$Z = 3$	
	Log likelihood	AIC	Log likelihood	AIC	Log likelihood	AIC
1	-119.36	240.74	-119.36	242.79	-119.36	244.86
2	-119.36	246.95	-119.36	251.21	-119.36	255.56
3	-119.36	251.21	-119.36	257.78	-119.36	264.59

With the estimated belief distribution function, we can calculate the percentage of people who consider the event may happen if there are no odds (or equal odds on both sides) in the prediction market, as illustrated in Figure A2. In all three datasets, the bettors' beliefs are more extreme than the actual event probability. If the event probability is less than 0.5, the final bet ratio will be lower than the event probability. If the event probability is higher than 0.5, the final bet ratio is higher than event probability.



Copyright of MIS Quarterly is the property of MIS Quarterly and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.