Corporate Predicton Markets: Designing an Information Aggregation Mechanism

Guide: Prof Bikramaditya Datta

Anshul Goel
15817110

Department of Economic Sciences
Indian Institute of Technology, Kanpur

June 9, 2020

Outline

Introduction: Prediction Markets (I)

- The conventional forecasting methods rely on application of statistics tools on the data and with the advent of Machine Learning and AI, the reliance on data has increased many folds
- But, there are many questions that prevail in the field of business, financial markets, policy making or consumer markets that require human intellect in interpreting the situation, figuring out possible scenarios and coming to a decision, all of which is extremely complex to model using statistical tools
- Consider a hypothetical situation: the defence minister of country A makes a statement against its enemy country B. Should country B go to war? It is the human intellect that caters to the various factors like timings of the declaration, the actual words used, the mood and strength of statement. All of this is difficult to capture by statistical models and would require huge data to just extrapolate the situation from past, which may not always be available

Introduction: Prediction Markets (II)

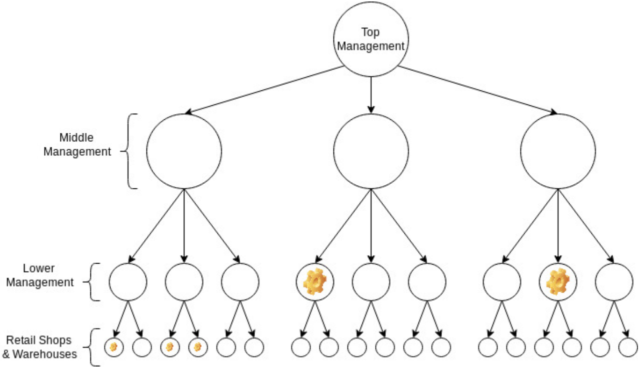
- This is what the prediction markets promise to deliver. As defined by Wikipedia, prediction markets are exchange-traded markets created for the purpose of trading the outcome of events
- Many studies have found that these markets have provided at par or better forecasts than the expert forecasters be it sports, politics or film industry
- Prediction markets collect their power from the idea of wisdom of crowds - a group of people can forecast better than an expert when certain conditions are met. These conditions are:
    1. diversity of opinion (each participant have some private information, even if it is an eccentric interpretation of the known facts)
    2. independence of opinion (participant's opinion not determined by opinion of others)
    3. decentralisation (participants can draw local knowledge)
    4. Aggregation (a mechanism to aggregate these judgements to a collective decision)

Introduction: Corporate Prediction Markets

- The prediction markets are used in diverse fields, some of which are listed here:
  1. Replication markets by DARPA to predict the replicability of studies in Social Science LINK
  2. A decentralised prediction market on almost all walks. be it Politics, Sports, culture. LINK
  3. Predicting the timing, nature and impact of scientific and technological advances and breakthroughs LINK
  4. Aggregation of people's information and using it in financial markets LINK
- Numerous corporations like Google, Microsoft, Motorola, JJ, Ford, HP, BestBuy have realised their strength and have created internal prediction markets as a method of information aggregation in the organisation. These internal markets are called corporate prediction markets

How it works? (I)

- The working of corporate prediction markets can be explained through the diagram below:



- The circles with icon represents the players with valuable information. These players generally don't have any incentive or mechanism to pass the information up the corporate ladder. Prediction markets aggregate this information to aid decision making

How it works? (II)

Below are some examples from BestBuy which created an internal corporate prediction called Tag-Trade:

1. Choosing between product categories: Circuit City, a competitor of BestBuy, was criticised for focusing on HD/DVD format than BlueRay. Tag-Trade helped Best Buy to decide on what product to focus with markets predicting a 77% market share for BlueRay compared to the actual 74% market share. BestBuy used these results and it dropped the HD/DVD format altogether.
2. Sales estimates: A security was traded which predicted quarterly sales of a computer-service package. Market's estimate was 33% lower than the official sales forecast with actual sales 25% lower. The package was later withdrawn.
3. Meeting the deadlines: The above package was redesigned and a market was run on the question if the relaunch meets its deadline. The market predicted it would with 85% probability, and the product was actually released on time.

Creating a Corporate Prediction Market: The Big Questions

There are some questions that needs to be answered for creating and running a prediction market:

1. What should be the incentive provided so that players answer truthfully and don't try to manipulate? Should it be monetary or non-monetary?
2. How to create 'thick' markets?
3. What should be the format of questions? Should the questions have multiple choice answers or should no option be provided and only a range is given? On the basis of it, what shoud be the scoring mechanism?
4. How to aggregate the information? Should methods like voting, stock markets, point spreads, pari-mutuel odds or future contracts be used?
5. What should be the target users for these markets?

Creating a Prediction Market in IIT Kanpur (I)

To answer these questions, we designed an internal prediction market for IIT Kanpur which provided valuable insights in designing prediction markets and choosing suitable market mechanism. The components of the prediction market has been discussed below:

- Questions: Prediction markets should only include questions for which some information is available to the players (in our case, to the IIT Kanpur residents). If there is no information available related to the question, players will just return garbage responses. Some possible questions:
  1. Questions related to the institute's technical fest (Techkriti) like the number of sponsors, will budget exceed the last year's budget, total budget for this year, total number of panelists, etc
  2. What are be the number of rolls sold by Kathi Rolls in campus in a day? (Can create a benchmark by asking the estimates of canteen owners, average them and compare with prediction market results)
  3. Some fun questions included to improve market's liquidity

Creating a Prediction Market in IIT Kanpur (II)

- Aggregation Mechanism: There are two possible methods that can be employed for aggregating information in our prediction markets:
  1. CDA (Continuous Double Auction): It requires two parties to bet against one another just like the stock market. Additional complexities can be introduced by adding order book feature and introducing different type of orders.
  2. AMM (Automated Market Maker): It allows the players to trade without the need of another party, directly with the market creator. It works on the principle of Logarithmic Market Scoring Rule (LMSR) which auto-adjusts price, number of shares based on a pre-decided parameter which limits the maximum loss of market creator and also determines the market depth
  3. Though the CDA market is simple and more intuitive to use, AMM suits best for our case due to its ability to work in low liquidity. Since these markets can't involve monetary payoff (discussed later), high continuous liquidity can't be expected and hence, AMM suits best.

Creating a Prediction Market in IIT Kanpur (III)

- Incentives: Some possible incentives can be as follows:
  1. Monetary: Monetary benefits may not be a suitable solution for our market owing to legal constraints. Another problem is that it leads to increase in bias as the players tend to place bets on low probability outcomes to gain higher payoff. To counter this, Google introduced a mechanism in which the people will certain number of tickets to a lottery as per their score. Thus, their score mattered more than the instant win
  2. Non-Monetary: The incentives should encourage people to participate. Also, they have to be decided on the basis of players' populace. Google used t-shirts as an incentive. In our case, the incentive currency can be some trendy t-shirts, internships/networking, career prospects, Mentoring, etc as most of our users will be undergraduate & postgraduate students
  3. Reputation: Reputation based incentives worked well for coporates like Ford. We can introduce the designations like 'superforecasters' for the top performers, motivated from the book `Superforecasting`

Creating a Prediction Market in IIT Kanpur (IV)

- Duration:
  1. Duration of prediction market for a question will be 1 week on the lines of HP's prediction market
  2. The markets can be run for around 4-5 hours or less a day to motivate people to check their current positions and change according to the current price. As the major changes will happen in that time, more transactions would occur leading to higher liquidity
- Software:
  1. There are some proprietary softwares for running prediction markets but are very expensive to buy
  2. There are some open source and free to use softwares too. One of the best in this segment is `Zocalo`, though setting it up would require some work with its creator. We connected with its creator Chris Hibbert who is willing to work on its update
  3. A better alternative for discrete choice questions can be `Prediki` which allows creating public or private prediction market at no or meager cost and a numerous other functionalities. A sample question created by me can be found here: `Prediki`

Creating a Prediction Market in IIT Kanpur (V)

- Issues/Limitations
  1. Software:A major issue with Prediki is that the exact model employed is not known. We connected with their team for the same but they just told that it works on LMSR's variant. Also one can't change the price very much by trading that may lead to strong bias towards prior
  2. Participation: The success of the market requires continuous participation. This problem is faced by all corporations that employ prediction markets and could be more severe for us as we can't introduce monetary benefits for the fear of resemblance to gambling
  3. Legal: The questions asked in the prediction market can be based on activities, policies or some future projects of IIT Kanpur which may lead to some legal bottlenecks, before or even after the market has ended. There is also a possibility of betting on inside information for questions of strategic importance to the institute. Hence, an extra level of care is required while deciding the questions

Market Design: Probabilistic v/s Non-Probabilistic

- One can make predictions about uncertain future events in terms of probabilities (Probabilistic Method) or Yes/No predictions (Non-Probabilistic Method)
- The primary concern is that the market price converges to the true probability
- From the perspective of a prediction market designer, there are different criteria to think about while comparing Probabilistic & Non-Probabilistic Method
- The project verifies the convergence of market price to true probability in both methods, and consider several versions like including the option of saving plus betting versus only betting and no saving
- This is verified by taking the classic betting example - horse race betting. I considered the parimutuel horse betting to check convergence in both methods. After that, I compared the 2 methods on other criteria and analysed the pros and cons of each method

Market Design: Probabilistic Method (I)

- Consider a 2 horse race. There are n players who are betting on either horse A or B. Each player has wealth $w_i, i = 1, ..., n$ and bets all his wealth on horses (no saving). The rationale behind the assumption of no saving is justified later
- The odds on horse A and B winning are o_A and o_B respectively
- Players can either bet on horse A or B alone or split their money between 2 horses by betting r fraction on A and 1 - r on B, where $0 \leq r \leq 1$. The player will maximise her expected utility. Say, a player has believes A will win & probability of belief is 'a' and b(=1-a) is the belief that B will win. Then that player will maximise his utility given his belief
- Expected Utility(EU) $= a \log(wro_A) + (1 - a) \log(w(1 - r)o_B)$
- Maximising the expected utility gives $r = a$. Thus, the player will bet his belief on each horse

Market Design: Probabilistic Method (II)

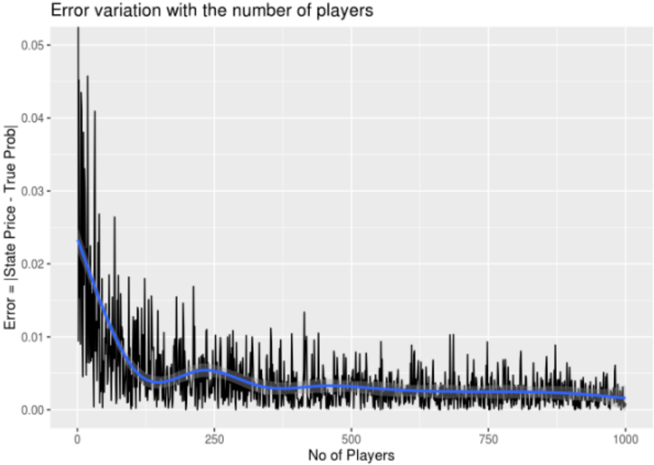
- Assume that racetracks charge no fees. So, the total amount bet is equal to total amount paid.
  For the racetrack:
  Total bet amount (W) $= w_1 + w_2 + ... + w_n$
  Total amount paid $= o_A \sum_{i=1}^{N} w_i a_i$
- Equating the two, we get: $o_A^{-1} = \sum_{i=1}^{N} f_i a_i$ & $o_B^{-1} = \sum_{i=1}^{N} f_i b_i$, where $f_i = \frac{w_i}{W}$
- Inverse odds (o_A^{-1} & o_B^{-1}) are called state prices as they are the price of a dollar in the event that a future state of the world (in our case, horse A or B wins) is reached
- It is easy to see that $o_A^{-1} + o_B^{-1} = 1$. You can bet $o_A{-}1$ on horse A and o_B^{-1} to get 1 unit back for sure. So, whatever amount one want to save, one can bet in these proportions. That's the reason why it was assumed that no player is saving

Market Design: Probabilistic Method (III)

- For verifying analytically if they converge, assume each player has equal wealth (=w) and the probability belief of each player is coming from a distribution centered around the true probability (p_{True}). Therefore, $o_A^{-1} = \frac{1}{n}\sum_{i=1}^{n} a_i$ becomes the average of the belief of all players
- Since, a_i is derived from a distribution centered around p_{True}, $o_A - 1$ should converge to p_{True} (from the Law of Large Numbers)
- We performed a simulation where the beliefs followed `PERT Distribution` with true probability of 0.6. We took the PERT distribution because we know mean value(=0.6), minimum value(=0) & maximum value(=1)
- We also performed the same exercise using Beta Distribution with parameters chosen such that mean becomes 0.6
- In both the cases, the state prices converge to true probability verifying the claim that in the probabilistic method, market price (state price) converges to true probability

Market Design: Probabilistic Method (IV)

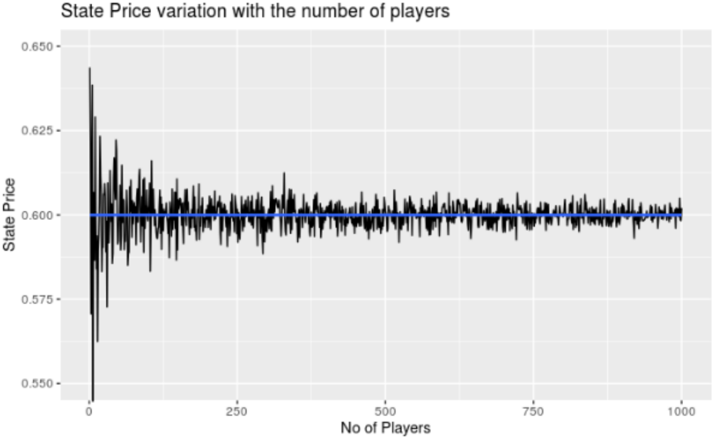
- For PERT distribution:



- Error which is the absolute difference between state price and true probability is converging to zero as number of players increase

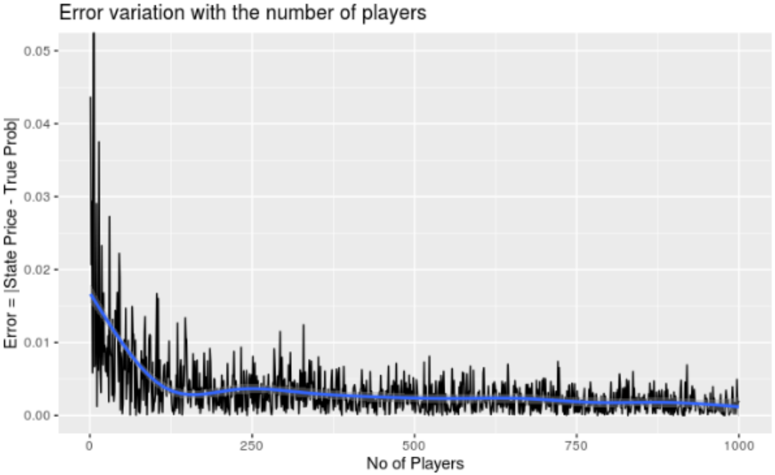
Market Design: Probabilistic Method (V)

- For PERT distribution:



- State price is converging to the true probability as the variations are dying out with increasing players

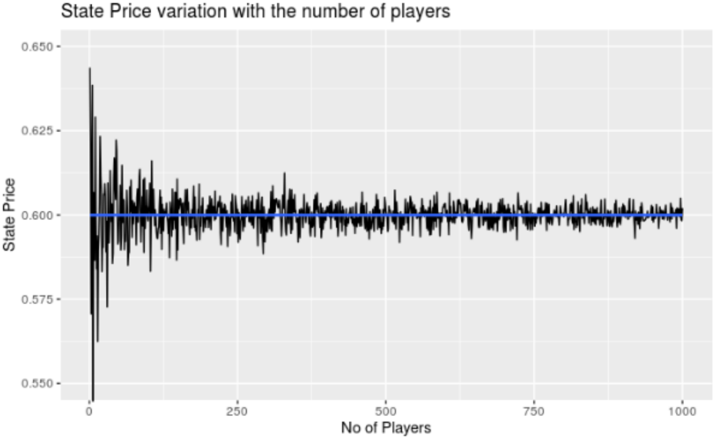
- For Beta distribution:



- Error which is the absolute difference between state price and true probability is converging to zero as number of players increase

Market Design: Probabilistic Method (VII)

- For Beta distribution:



- State price is converging to the true probability as the variations are dying out with increasing players

Market Design: Non-Probabilistic Method

- As the case with the probabilistic method, we assume 2 horse race with each player having wealth w but now the player can only bet on 1 horse, either A or B
- We introduce the signalling framework where the players receive either a High signal which means horse A is more likely to win and Low signal which means horse B is a better bet
- As in the previous case, racetracks are just redistributing the wealth. So, the total bet amount is equal to the total amount awarded
- Thus, ${o_A}^{-1} = \sum_{i=1}^{N} a_i \frac{w_i}{W}$ and ${o_B}^{-1} = \sum_{i=1}^{N} b_i \frac{w_i}{W}$
- $a_i = \begin{cases} 1, \text{if High Signal is received} \\ 0, \text{if Low signal is received} \end{cases}$
- Modelling a_i:
  - Each player has some prior probability(p) about each horse winning which is the public information. Thus, P(A win) = p & P(B win) = 1-p
  - We will assume that $p = \frac{1}{2}$ for sake of simplicity. It also means that both horses are considered to have equal chances of win publicly

Market Design: Non-Probabilistic Method (II)

- Modelling a_i (continued)
  - Signal: Given the best horse, a signal is generated (high or low) with probability q($>$ 0.5)
    Pr(H signal is observed | A win) = q (P(H | A)=q)
    Pr(L signal is observed | B win) = q (P(L|B)=q)
- Posterior: Players will bet after they received the signal (H or L), keeping in mind the signal and public information (prior)
- P(A win | H signal received) = $\frac{pq}{pq+(1-p)(1-q)} = q$ (Since p = 0.5)
- Similarly, it can be shown that P(B | L)=q, P(A | L)=1-q, P(B|H)=1-q
- Therefore, $o_A^{-1} = \frac{\sum_{i=1}^{n} a_i}{n} = \frac{\text{Number of times H signal observed}}{\text{Total Signals received (total number of participants)}}$
- Thus, o_A^{-1} = probability of receiving a high signal which is equal to p
- Hence, the market price converges to true probability

Market Design: Non-Probabilistic Method (III)

- Now, we will perform the calculations for the case where savings is allowed. All the calculations will be done considering that High signal is observed but will remain same for the Low signal
- Each player will bet α fraction of her wealth on horses and save $1 - \alpha$ fraction
- Remaining framework will be same to the previous model with no saving
- Expected Utility in this case would be:
  Pr(Horse A win)log(Wealth received if A win) + Pr(Horse B win)log(Wealth received if B win)
- $EU = q \log(w(1-\alpha) + w\alpha o_A) + (1-q) \log(w(1-\alpha))$
- Maximising the expected utility gives $\alpha = \frac{qo_A - 1}{o_A - 1}$, as $\alpha > 0 and \alpha < 1$, this gives $qo_A > 1$ & $q < 1$

Market Design: Non-Probabilistic Method (IV)

- For racetrack, total wealth $= \sum_{i=1}^{n} w_i \alpha_i$, total payoff for players who bet on A $= \sum_{i=1}^{n} w_i \alpha_i a_i o_A$
- Equating the 2 gives:
  $o_A^{-1} = \frac{n_2}{n_1+n_2} + \frac{n_1-n_2}{n_1+n_2} q$
  n_1 means the number of players who bet on horse A
  n_2 are the number of players who bet on horse B
- Substituting the value in the expression gives $\alpha = \frac{n_2(2q-1)}{n_1+(n_2-n_1)q}$.
  Here the conditions on α gives $q > 0.5 \quad q < 1$ which already satisfies the past conditions

Market Design: Non-Probabilistic Method (V)

- Interpretation:
  1. $q = 0.5$, $o_A^{-1} = 0.5$ and $\alpha = 0$
     Interpretation: If it is uncertain which horse is more likely to win (q=0.5), the player has no benefit of betting and is better-off by just saving all the amount and get EU = log(w)
  2. q = 1, $o_A^{-1} = \frac{n_1}{n_1+n_2}$ and $\alpha = 1$
     Interpretation: If q = 1, i.e. player is sure about the victory of the horse then the player is better-off by betting all his wealth on horse and get expected payoff of log(wo_A)
  3. $0.5 < q < 1$, $o_A^{-1} = \frac{n_2}{n_1+n_2} + \frac{n_1-n_2}{n_1+n_2} q$ and $\alpha = f(n_1, n_2, q)$
     Interpretation: Player will first check between the utility gained from betting and not betting and bet only if the expected utility is higher

Comparison Between the Methods

4 criteria have been considered to compare between Probabilistic & Non-Probabilistic Method:

- Favourite-Longshot bias:
  1. Several studies have highlighted the prevailing bias in probabilistic setup. Players tend to over bet on underdogs and under bet on favourites. In some cases, like in Google's internal prediction market, reverse Favourite-Longshot bias was observed
  2. In the Probabilistic setup, $o_A^{-1} = \frac{\sum_{i=1}^{n} a_i}{n}$, i.e. the state prices is equal to each player's beliefs as well as bias. Thus, the bias is incorporated in the state prices (prediction market price)
  3. In the case of Non-Probabilistic setup, state price is equal to the fraction of players who received the signal. As the players can only put all their wealth on one option and it is not rational to go against own information. Thus, the inherent bias doesn't play a significant role
  4. Hence, we should prefer Non-Probabilistic method to overcome this factor

Comparison Between the Models

- Under Pricing of Extreme Outcomes:
  1. This refers to the problems where the outcomes have a relative ranking like the feedback of a product: Very Poor(1), Poor(2), Average(3), Good(4), Very Good(5). A player would avoid too much weightage on extreme outcomes and would even put a lower weightage than their belief due to extreme aversion behaviour whereas in Non-Probabilistic setup player has to bet on a single option and would do on the option she seems most likely rather than falling on the cushion of moderate values
  2. In case of probabilistic setup, players avoid betting on extremes which under price the extremes and overprice the less extreme outcome
  3. Players undercut their bet on extremes in favour to put a share of wealth in moderates. However, in case of a non-probabilistic setup player can only bet on one outcome which will require a rational player to bet on the outcome she deem most probable
  4. It will also push the players to exert effort in reinforcing their beliefs as all of their wealth is put on one outcome
  5. Thus, Non-Probabilistic method would be preferred here

Comparison Between the Models

- Participation
  1. In case of non-probabilistic setup, players can bet on one of the options only and not mitigate the risk by betting some fraction of other outcomes to, as per their probability beliefs
  2. This might cause reduced participation, especially, in the real money markets.
  3. If we consider the past criteria of pricing of extreme outcomes, the players won't even participate if they don't have sufficient information or strong belief for one of the option
  4. Even in binary outcomes, probabilistic methods provide a method to maximise their utility by betting on both outcomes (refer expected utility functions we employed in the past slides). However, in case of close call between options say 45% for option A and 55% for option B, it might cause to avoid the market altogether, especially when their money is at stake and would be better-off by saving ($\alpha = 0$)
  5. Participation is a critical factor in any prediction market, especially the corporate prediction markets. Thus, we have a trade off here between participation and bias

Comparison Between the Models

- Overall Information Aggregation:
  1. Though non-probabilistic setup is helpful in reducing the effect of bias, it also leads to loss in information
  2. In a non-probabilistic setup, we are only aggregating the knowledge of people who believe the outcome is most likely and discarding his information about other outcomes. This effect is more stringent in the market with multiple outcomes
  3. Whereas in case of probabilistic setup, the player's belief about each outcome is considered in aggregating the odds and hence incorporates more information
  4. Consider an example where player believes that there is a 5% chance that horse A will win and 95% chance that horse B will win. In probabilistic setup, the information of both 95% and 5% would have been incorporated into the state price in o_A^{-1} & o_B^{-1} whereas in the case of non-probabilistic setup, only information that is conveyed is that horse A is more likely to win than horse B win and hence leads to loss of information
  5. Again, there is a trade-off between information volume and bias

Acknowledgement

I express my sincere gratitude to my advisor, Prof Bikramaditya Datta who helped me throughout this project and provided me required resources, expert council and a proper framework as the work progressed. This project would not have been possible without him.

I would also like to acknowledge the role of my friends and family who helped me directly and indirectly in the completion of my project.

Last but not the least, to my department and institute for providing me the opportunity to pursue this project and providing adequate infrastructure and support

Reference

- Surowiecki, James. The Wisdom of Crowds. New York: Anchor Books, 2005.
- Snowberg E, Wolfers J & Eric Zitzewitz. Prediction Markets for Economic Forecasting. CESifo Working Paper Series 3884, CESifo, 2012.
- Cowgill Bo & Eric Zitzewitz. Corporate Prediction Markets: Evidence from Google, Ford and Firm X*. Review of Economic Studies, 2015.
- Rothschild D & Rajiv Sethi. Trading Strategies and Market Microstructure: Evidence from a Prediction Market. Microsoft Research, 2015
- Aldous, David. On Chance and Unpredictability: Lectures on the Links Between Mathematical Probability and the Real World.
- Wolfers J & Eric Zitzewitz. Interpreting Prediction Market Prices as Probabilities

References

- Cowgill B, Wolfers J & Eric Zitzewitz. Using Prediction Markets to Track Information Flows: Evidence from Google.
- Roth, Alvin. Who Gets What and Why. 2016
- Easley D & Jon Kleinberg. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press, 2010
- Montgomery T, Steig P, Cavaretta M & Paul Moraal. Experience from Hosting a Corporate Prediction Market: Benefits beyond the Forecasts. ACM 978-1-4503-2174-7/13/08
- Borison A & Gregory Hamm. Prediction Markets: A New Tool for Strategic Decision Making. California Management Review, Summer, 2020
- Jacobs, Vincent. Prediction Markets: How They Work & How Well They Work. 2009

References

Bottazzi G D. Giachini (2019): Far from the madding crowd: collective wisdom in prediction markets, Quantitative Finance, DOI: 10.1080/14697688.2019.1622285

Ali, Mukhtar. Probability and Utility Estimates for Racetrack Bettors. Journal of Political Economy, 1977