Crowd wisdom and prediction markets

Yanwei Jia

National University of Singapore

Joint works with
Min Dai (NUS) and Steven Kou (BostonU)

Overview

1. Introduction
2. Review of Bayesian truth serum
3. Design of a prediction market
4. Main results
   - Hybrid Market-Survey Estimator
   - Adjusted Market Estimator
5. Numerical results
   - Hybrid estimator
   - Adjusted market estimator
6. Conclusion

Content

1 Introduction

2 Review of Bayesian truth serum

3 Design of a prediction market

4 Main results
- Hybrid Market-Survey Estimator
- Adjusted Market Estimator

5 Numerical results
- Hybrid estimator
- Adjusted market estimator

6 Conclusion

A motivating example

- Philadelphia is the capital of Pennsylvania, yes or no?
- The majority opinion can be wrong, as shown in Prelec et al. (2017).
  - Most people vote yes;
  - Confidences associated with yes and no voters are roughly similar;
  - Respondents voting no expect to be minority, while respondents voting yes believe that most people agree with them.

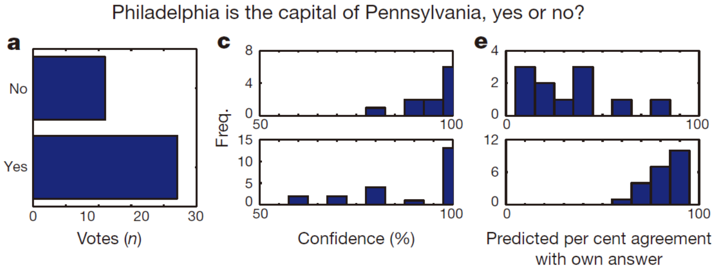


Figure 1: Survey results in Prelec et al. (2017). a) shows the number of respondents answering yes and no. c) shows confidence about their answer being correct. e) shows their predicted percentage of others who agree with their answer. The figure is taken from Prelec et al. (2017).

Motivation

- Prelec (2004) introduces an innovative survey design called the Bayesian truth serum (BTS) by asking two questions and designs the mechanism to encourage truth-telling.
- Prelec et al. (2017) propose a consistent estimator that selects the best answer based on survey questions under the BTS framework.

Motivation

- An important area of Fintech is crowd wisdom, which aims at using the collective opinion of a group to make predictions.
- Prediction markets gain more attention thanks to the digital innovation (significantly reduces costs in designing and participating in prediction markets).
- Is there a consistent estimator based on prediction markets?

Our contribution

- We complements Prelec et al. (2017) by proposing two estimators that are consistent under the BTS settings but with different regularity conditions.
- One market-adjusted estimator based solely on market inputs and a hybrid estimator based on both market and inputs from one survey question (instead of two questions).

How to discover truth

Approach	Pros	Cons
Ask experts	easy to implement	biased, difficult to identify experts
Poll	crowd wisdom, more than one questions	lack of incentives, non-response bias, costly
Prediction markets	crowd wisdom, many participants, real-money incentives, instant update	observable and verifiable events, market frictions
Bayesian/experimental markets	crowd wisdom, non-verifiable questions, incentive compatible	difficult to design, costly to maintain

Literature review on BTS

Literature	Objectives
Prelec (2004)	Show the truth-telling is the equilibrium outcome to maximize BTS score
Baillon (2017)	Design a Bayesian market to show truth-telling is the equilibrium outcome for a binary question
Prelec et al. (2017)	Propose an estimator that can deduce the objective truth under certain conditions
This paper	Propose two estimators that can deduce the objective truth under different conditions

Comparison of literature on prediction markets

Table 1: **A summary of analyses on equilibrium prediction market prices.** WZ (2006) means Wolfers and Zitzewitz (2006), and OS (2015) means Ottaviani and Sørensen (2015).

Literature	Total number of the agents	Utility function	Risk aversion	Wealth
WZ (2006)	continuum	quadratic, HARA,	homogeneous	heterogeneous
OS (2015)	continuum	risk neutral, CARA, CRRA	heterogeneous	heterogeneous
This paper	finite	CARA, CRRA, risk neutral	heterogeneous	heterogeneous
Main results				
WZ (2006)	(unadjusted) Market prices correspond with average beliefs under certain conditions.			
OS (2015)	(unadjusted) Market prices underreact to information under certain conditions.			
This paper	Give adjusted market prices as consistent estimators under the BTS framework.			

A comparison of different main assumptions for the consistency

(a) Panel A: Binary outcome

	Assumptions						
	Prelec et al. (2017)				This paper		
	A1	A2	A3	A4	A5	A6	A7
BTS (Prelec et al., 2017)	✓	✓		✓			
Market-survey hybrid BTS	✓	✓	✓				
Adjusted market BTS					✓	✓	✓

(b) Panel B: Multiple outcome

	Assumptions						
	Prelec et al. (2017)				This paper		
	A1	A2	A3	A4	A5	A6	A7
BTS (Prelec et al., 2017)	✓	✓	✓	✓			
Adjusted market BTS					✓	✓	✓

Content

1 Introduction

2 Review of Bayesian truth serum

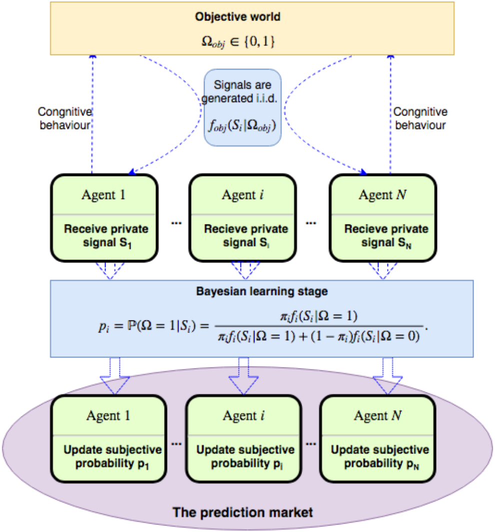
3 Design of a prediction market

4 Main results
- Hybrid Market-Survey Estimator
- Adjusted Market Estimator

5 Numerical results
- Hybrid estimator
- Adjusted market estimator

6 Conclusion

Cognitive and decision making process



Assumptions in Prelec et al. (2017)

Assumption A1

All agents agree on a common prior distribution $\pi^k = \mathbb{P}(\Omega = k) \in (0, 1)$.

Assumption A2

All agents agree on the subjective likelihood function, which is also equal to the objective likelihood function.

- Assumption A1 and A2 implies that agents are identical except for their private information;
- Assumption A1 and A2 are necessary in most literature on BTS.

Assumptions in Prelec et al. (2017)

Assumption A3

$\mathbb{P}(\Omega = k | S_i = k) > \mathbb{P}(\Omega = k | S_i = j)$, *for any* $j \neq k$.

Assumption A4

$\mathbb{P}(\Omega = k | S_i = k) > \mathbb{P}(\Omega = j | S_i = k)$, *for any* $j \neq k$.

- Assumption A3 is the most crucial assumption for the consistency in Prelec et al. (2017);
- Assumption A4 implies that agents with $S_i = k$ will believe $\Omega = k$ is more likely the correct answer;
- In the case of binary outcome event, Assumption A3 is implied by Assumption A4 since

$$\mathbb{P}(\Omega = 1 | S_i = 1) > 0.5 > \mathbb{P}(\Omega = 0 | S_i = 1),$$

$$\mathbb{P}(\Omega = 0 | S_i = 0) > 0.5 > \mathbb{P}(\Omega = 1 | S_i = 0).$$

Review of the estimator in Prelec et al. (2017)

- Question 1, "which one is more likely to happen, $\Omega = 1$ or $\Omega = 0$", and the answer is denoted by v_i.
    - Estimating $\mathbb{P}(S_i = k|\Omega_{obj})$.
- Question 2, "what is the proportion of the population who will answer $\Omega = k$", and the answer is denoted by ξ_i^k.
    - Estimating $\mathbb{P}(S_j = k|S_i)$.
- The estimator is constructed by

$$\mathbf{1}_{\left\{\frac{\sharp\{i:v_i=1\}}{m_{01}} > \frac{\sharp\{i:v_i=0\}}{m_{10}}\right\}}, \tag{1}$$

where $m_{kl} = \dfrac{1}{\sharp\{i : v_i = k\}} \displaystyle\sum_{i:v_i=k} \xi_i^l$.

- $M_{kl}\mathbb{P}(S_i = k) = M_{lk}\mathbb{P}(S_i = l)$, where $M_{kl} = \mathbb{P}(S_j = l|S_i = k)$.
- $\mathbb{P}(\Omega = \Omega_{obj}|S_i = k) \propto \frac{\mathbb{P}(S_i=k|\Omega_{obj})}{\mathbb{P}(S_i=k)}$.

Truth-telling mechanism in Prelec (2004)

- The payoff to respondent i is

$$\sum_k \mathbf{1}_{\{v_i=k\}} \log \frac{\bar{v}^k}{\bar{\xi}^k} + \alpha \sum_k \bar{v}^k \log \frac{\xi_i^k}{\bar{v}^k},$$

where $\bar{v}^k = \frac{\sharp\{i:v_i=k\}}{N}$, $\log \bar{\xi}^k = \frac{1}{N}\sum_{i=1}^N \log \xi_i^k$ and $\alpha > 0$.

- The truth-telling is a Nash equilibrium, with

$$v_i = S_i,\ \xi_i^k = \mathbb{P}(S_j = k | S_i).$$

Content

1 Introduction

2 Review of Bayesian truth serum

3 Design of a prediction market

4 Main results
- Hybrid Market-Survey Estimator
- Adjusted Market Estimator

5 Numerical results
- Hybrid estimator
- Adjusted market estimator

6 Conclusion

Design of a prediction market

- Two securities are traded, each of which bets on the outcome of a toss of a coin with payoff 1 dollar, denoted by H and T.
- Given the market price $\bar{p}$ for H and $\bar{q}$ for T, for agent i, he solves the following expected utility maximization problem

$$\max_{\bar{p}x+\bar{q}y=w_i} p_i U_i(x) + q_i U_i(y), \tag{2}$$

where w_i is the initial wealth he will invest in this market, (x, y) are the number of shares in each security and subjective probability (p_i, q_i), U_i is a concave utility function.

Assumption A5

The agents in the economy either all have constant relative risk aversion (CRRA) preferences or constant absolute risk aversion (CARA) preferences, where the risk averse coefficients $\gamma_i > 0$ *are independent identically distributed (i.i.d.), are independent of signals and prior probabilities, and satisfy* $0 < \mathbb{E}_{obj}[\frac{1}{\gamma_i^2}] < \infty$. *Initial wealth* $w_i > 0$ *are i.i.d. and independent of signals, prior probabilities, and risk aversion coefficients.*

Regularity Condition 1

$0 < \mathbb{E}_{obj}[w_i] < \infty$.

Equilibrium market price and optimal strategy

Proposition 1

(a) *If agent i has utility function $U_i(c) = -\frac{1}{\gamma_i} e^{-\gamma_i c}$, for $1 \leq i \leq N$, then the equilibrium price $\bar{p}$ exists and is the solution to*

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\gamma_i} \log \frac{p_i}{1-p_i} = \log \frac{\bar{p}}{1-\bar{p}} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\gamma_i}, \tag{3}$$

and optimal strategy (x_i, y_i) for agent i satisfies $x_i - y_i = \frac{1}{\gamma_i}(\log \frac{p_i}{1-p_i} - \log \frac{\bar{p}}{1-\bar{p}})$.

(b) *If agent i has utility function $U_i(c) = \frac{c^{1-\gamma_i}}{1-\gamma_i}$ and initial wealth w_i, for $1 \leq i \leq N$, then the equilibrium price $\bar{p}$ exists and solves the following*

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1 - \left(\frac{\bar{p}(1-p_i)}{(1-\bar{p})p_i}\right)^{\frac{1}{\gamma_i}}}{\left[\left(\frac{\bar{p}(1-p_i)}{(1-\bar{p})p_i}\right)^{\frac{1}{\gamma_i}} - 1\right]\bar{q} + 1} w_i = 0, \tag{4}$$

and optimal strategy (x_i, y_i) for agent i satisfies $\log x_i - \log y_i = \frac{1}{\gamma_i}(\log \frac{p_i}{1-p_i} - \log \frac{\bar{p}}{1-\bar{p}})$.

Content

1 Introduction

2 Review of Bayesian truth serum

3 Design of a prediction market

4 Main results
- Hybrid Market-Survey Estimator
- Adjusted Market Estimator

5 Numerical results
- Hybrid estimator
- Adjusted market estimator

6 Conclusion

A Hybrid Market-Survey Estimator

Definition 1

The market-survey hybrid estimator is defined as

$$\hat{\Omega}_{BTS,hybrid} = \mathbf{1}_{\{\tilde{p}_1 > \hat{\mu}_1\}}, \tag{5}$$

where $\tilde{p}_1 = \frac{\sharp\{i:x_i>y_i\}}{N}$, $\hat{\mu}_1 = \frac{\hat{m}_{01}}{\hat{m}_{10}+\hat{m}_{01}}$, *and* $\hat{m}_{01}$, $\hat{m}_{10}$ *can be calculated by*

$$\hat{m}_{10} = \frac{1}{\sharp\{i : x_i > y_i\}} \sum_{i:x_i>y_i} \xi_i^0, \quad \hat{m}_{01} = \frac{1}{\sharp\{i : x_i < y_i\}} \sum_{i:x_i<y_i} \xi_i^1.$$

Here, $\xi_i^1(\xi_i^0)$ *is agent* i*'s answer to the question "what is the proportion of the rest of the population who will invest more in* $H(T)$*".*

Consistency

Proposition 2

Under Assumptions A1, A2, A3 and Regularity Condition 1, $\hat{\Omega}_{BTS,hybrid} \to \Omega_{obj}$ a.s. $\mathbb{P}_{obj}$, i.e. the market-survey hybrid BTS estimator given in (5) is consistent.

- The position (x_i, y_i) is analogy of the first survey question;
- The survey question is equivalent to the second question;
- Why Assumption A4 can be relaxed?
    - Assumption A4 $\Rightarrow \{i : v_i = 1\} = \{i : S_i = 1\}$;
    - Assumption A3 + market clearing condition $\Rightarrow \{i : x_i > y_i\} = \{i : S_i = 1\}$;
    - $\{i : S_i = 1\}$ can still be identified so that $\mathbb{P}(S_j = k | S_i = 1)$ can be estimated consistently.
- Why Assumption A5 is not necessary?
    - we only require that agents with $p_i > \bar{p}$ will choose $x_i > y_i$.

An Adjusted Market Estimator

Assumption A6

The unobserved prior probability $\pi_i^k := \mathbb{P}_i(\Omega = k) \in (0, 1)$ *is drawn independently from a certain distribution, such that* $\mathbb{E}_{obj}[\log \pi_i^k] = \log \bar{\pi}^k$ *exists, for any* k*, and the prior probability does not depend on the risk aversion coefficient and initial wealth.*

- prior can be heterogeneous;
- relaxation of Assumption A1.

Regularity Condition 2

There exists $0 < \alpha < \frac{1}{2}$, such that subjective likelihood functions for all agents satisfy

$$||\log f_i(\cdot|\Omega = \Omega_{obj}) - \log f_{obj}(\cdot|\Omega_{obj})||_{L^2(\mathbb{P}_{obj})} = O(i^{\frac{1}{2}-\alpha}), \tag{6}$$

$$\frac{1}{N}\sum_{i=1}^{N} \log f_i(\cdot|\Omega = \Omega_{obj}) \to \log f_{obj}(\cdot|\Omega_{obj}) \textit{ in } L^1(\mathbb{P}_{obj}), \tag{7}$$

and there exists a probability density function (called counterfactual likelihood) $g(\cdot|\Omega^c_{obj}) \neq f_{obj}(\cdot|\Omega_{obj})$ for any event $\Omega^c_{obj} \neq \Omega_{obj}$, such that

$$||\log f_i(\cdot|\Omega = \Omega^c_{obj}) - \log g(\cdot|\Omega^c_{obj})||_{L^2(\mathbb{P}_{obj})} = O(i^{\frac{1}{2}-\alpha}), \tag{8}$$

$$\frac{1}{N}\sum_{i=1}^{N} \log f_i(\cdot|\Omega = \Omega^c_{obj}) \to \log g(\cdot|\Omega^c_{obj}) \textit{ in } L^1(\mathbb{P}_{obj}). \tag{9}$$

Assumption A7

The regularity condition 2 holds, and

$$\log \frac{\bar{\pi}^{\Omega_{obj}}}{\bar{\pi}^{j}} > -D_{KL}(f_{obj}(\cdot|\Omega_{obj}),\ g(\cdot|\Omega^{c}_{obj} = j)),$$

for any j, where $D_{KL}(f, g)$ is the Kullback-Leibler divergence between two distributions. Here $\bar{\pi}^{\Omega_{obj}}$ means that $\bar{\pi}^{\Omega_{obj}} = \bar{\pi}^{k}$ on the event $\Omega_{obj} = k$.

- likelihood functions can be heterogeneous but on average should be close to objective likelihood function;
- relaxation of Assumption A2.
- When $\bar{\pi}^{\Omega_{obj}} \geq \bar{\pi}^{j}$, Assumption A7 holds true automatically;
- $\bar{\pi}^{\Omega_{obj}}$ cannot be too small.

Definition 2

The adjusted market estimator is defined as

$$\hat{\Omega}_{obj,market} = \mathbf{1}_{\{\hat{p}_N > \frac{1}{2}\}},$$

and $\hat{p}_N$ *can be computed via*

$$\frac{1}{N} \log \frac{\hat{p}_N}{1 - \hat{p}_N} = \log \frac{\bar{p}}{1 - \bar{p}} + \frac{\bar{\gamma}}{N} \sum_{i=1}^{N} z_i, \tag{10}$$

where $\sum_{i=1}^{N} z_i = 0$ *for CARA utility,*
$\sum_{i=1}^{N} z_i = \sum_{i=1}^{N} (\log x_i - \log y_i)$ *for CRRA utility,*

$$\bar{\gamma} := \frac{1}{\mathbb{E}_{obj}[\frac{1}{\gamma_i}]}$$

.

Consistency

- market price may be dominated by wealthy investors;
- market price may be dominated by low risk averse investors;
- the wealth effect and the risk aversion effect get cancelled in the case of CARA utility but not for CRRA utility.
- The estimator is based on $\mathbb{P}(\Omega = 1 | S_1, \cdots, S_N)$.
- Asymptotically equivalent to MLE $\max_k \prod_{i=1}^{N} f_{obj}(S_i | \Omega = k)$.

Theorem 1

Suppose that Assumptions A5, A6, and A7 hold. Then the adjusted market estimator using (10) is consistent, i.e. $\hat{\Omega}_{obj,market} \to \Omega_{obj}$, $\mathbb{P}_{obj}$-a.s.

Content

1. Introduction
2. Review of Bayesian truth serum
3. Design of a prediction market
4. Main results
   - Hybrid Market-Survey Estimator
   - Adjusted Market Estimator
5. Numerical results
   - Hybrid estimator
   - Adjusted market estimator
6. Conclusion

Hybrid estimator vs BTS: binary case

Assumptions		Consistency of Estimator	
Prelec et al. (2017)	This paper	Prelec et al. (2017)	This paper
A4	A5	survey only	hybrid market-survey
✓	✓	✓	✓
✗	✓	✗	✓
✓	✗	✓	✓

Hybrid estimator vs BTS: binary case

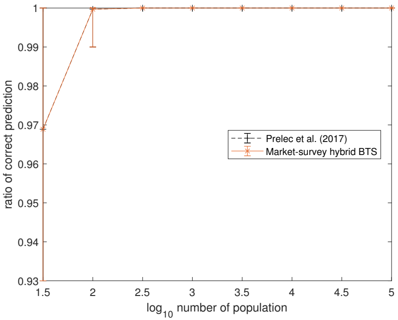


Figure 2: **The prediction performance when all assumptions are satisfied**. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. The question is whether Philadelphia is the capital of Pennsylvania, and the correct answer is no, i.e. $\Omega_{obj} = 0$. The parameters are taken from the Philadelphia problem in Prelec et al. (2017). $\pi = (\frac{5}{12}, \frac{7}{12})$ and the objective likelihood $f(1|1) = \frac{20}{21}$, $f(1|0) = \frac{2}{3}$. Agents are assumed to have CRRA utility with $\gamma_i \sim Unif(0.1, 0.5)$ and initial wealth $w_i \sim Unif(0, 10)$. In the simulation. $M = 100$ times and $B = 100$ times. The hybrid market-survey estimator appears to perform equally well compared with the survey only estimator in Prelec et al. (2017).

Hybrid estimator vs BTS: binary case

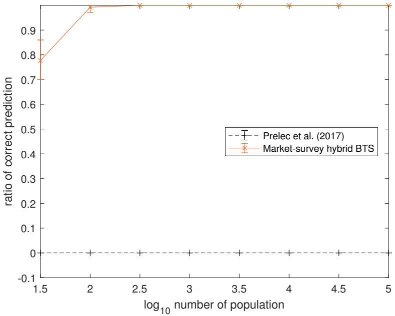


Figure 2: **The prediction performance when Assumption A4 is invalid**. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. $\Omega_{obj} = 1$, the prior $\pi = (0.6, 0.4)$ and the objective likelihood $f(1|1) = \frac{20}{21}$, $f(1|0) = \frac{2}{3}$. $P_{11} = 0.4878$, $P_{01} = 0.087$. Agents are assumed to have CRRA utility with $\gamma_i \sim Unif(0.1, 0.5)$ and initial wealth $w_i \sim Unif(0, 10)$. In this simulation, $M = 100$ and $B = 100$. It appears that in this case the survey only estimator is not consistent, while the hybrid estimator may still be consistent.

Hybrid estimator vs BTS: binary case

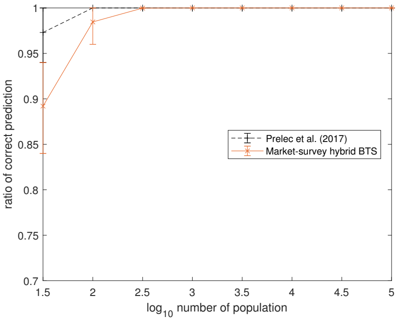


Figure 2: **The prediction performance when Assumption A5 is invalid**. In this case $\bar{\gamma} = 0$, because $\mathbb{E}_{obj}[\frac{1}{\gamma_i}] = \infty$. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. The question is whether Philadelphia is the capital of Pennsylvania, and the correct answer is no, i.e. $\Omega_{obj} = 0$. The parameters are taken from the Philadelphia problem in Prelec et al. (2017). $\pi = (\frac{5}{12}, \frac{7}{12})$ and the objective likelihood $f(1|1) = \frac{20}{21}$, $f(1|0) = \frac{2}{3}$. Agents are assumed to have CRRA utility with $\gamma_i \sim Exp(1)$ and initial wealth $w_i \sim Unif(0, 10)$. In the simulation, $M = 100$ and $B = 100$. Both estimators converge to the correct answer, even though Assumption A5 is invalid.

Adjusted market estimator vs BTS

(a) Panel A: Binary outcome case

Assumptions				Consistency of Estimator	
Prelec et al. (2017)		This paper		Prelec et al. (2017)	This paper
A1	A4	A5	A7	BTS	Adjusted market
✓	✓	✓	✓	✓	✓
✗	n/a	✓	✓	✗	✓
✓	✗	✓	✓	✗	✓
✓	✓	✗	✓	✓	✗
✓	✓	✓	✗	✓	✗

(b) Panel B: Multiple outcome case

Assumptions					Consistency of Estimator	
Prelec et al. (2017)			This paper		Prelec et al. (2017)	This paper
A1	A3	A4	A5	A7	BTS	Adjusted market
✓	✓	✓	✓	✓	✓	✓
✗	n/a	n/a	✓	✓	✗	✓
✓	✗	✓	✓	✓	✗	✓
✓	✓	✗	✓	✓	✗	✓
✓	✓	✓	✗	✓	✓	✗
✓	✓	✓	✓	✗	✓	✗

Adjusted market estimator vs BTS: binary case

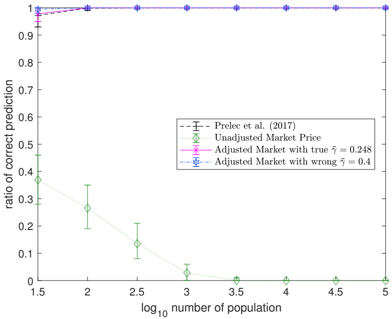


Figure 3: **The prediction performance when all assumptions are satisfied**. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. $\Omega_{obj} = 0$, with the prior $\pi_i = (0.5, 0.5)$ and the objective likelihood $f(1|1) = \frac{20}{21}$, $f(1|0) = \frac{2}{3}$. Agents are assumed to have CRRA utility with $\gamma_i \sim Unif(0.1, 0.5)$, and initial wealth $w_i \sim Unif(0, 10)$. In the simulation, $M = 100$ and $B = 100$. Both estimators are consistent when all assumptions hold true. And our market adjusted estimator is also robust to risk aversion parameter.

Adjusted market estimator vs BTS: binary case

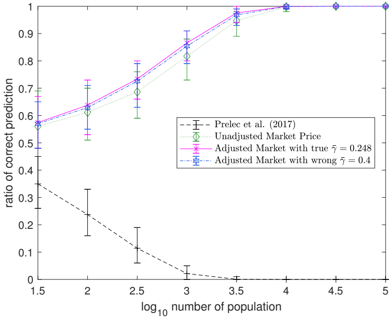


Figure 3: **The prediction performance when Assumption A1 is invalid**. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. $\Omega_{obj} = 1$, with the prior $\pi_i^1 \sim \mathit{Unif}(0.3, 0.7)$ and the objective likelihood $f(1|1) = 0.5$, $f(1|0) = 0.6$. Agents are assumed to have CRRA utility with $\gamma_i \sim \mathit{Unif}(0.1, 0.5)$, and initial wealth $w_i \sim \mathit{Unif}(0, 10)$. In the simulation, $M = 100$ and $B = 100$. When prior is heterogeneous, BTS may be inconsistent while our estimator still appears to be consistent.

Adjusted market estimator vs BTS: binary case

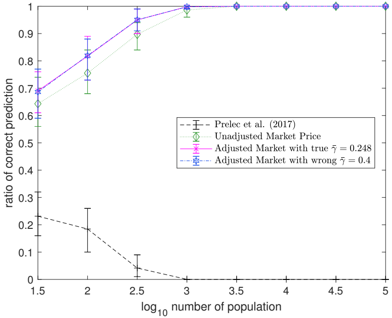


Figure 3: **The prediction performance when Assumption A4 is invalid**. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. $\Omega_{obj} = 1$, with the prior $\pi_i^1 = 0.5$ and the objective likelihood $f(1|1) = 0.5$, $f(1|0) = 0.6$. Agents are assumed to have CRRA utility function with risk aversion coefficient $\gamma_i \sim Unif(0.1, 0.5)$, and initial wealth $w_i \sim Unif(0, 10)$. In the simulation, $M = 100$ and $B = 100$. Assumption A4 is invalid in this case, but our estimator still leads to the correct answer.

Adjusted market estimator vs BTS: binary case

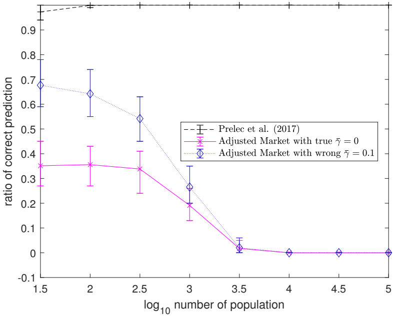


Figure 3: **The prediction performance when Assumption A5 is invalid**. In this case $\bar{\gamma} = 0$, because $\mathbb{E}_{obj}[\frac{1}{\gamma_i}] = \infty$. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. $\Omega_{obj} = 0$, with the prior $\pi_i = (0.5, 0.5)$ and the objective likelihood $f(1|1) = \frac{20}{21}$, $f(1|0) = \frac{2}{3}$. Agents are assumed to have CRRA utility with $\gamma_i \sim Exp(1)$ and initial wealth $w_i \sim Unif(0, 10)$. In the simulation, $M = 100$ and $B = 100$. Our estimator can be inconsistent if Assumption A5 is invalid.

Adjusted market estimator vs BTS: binary case

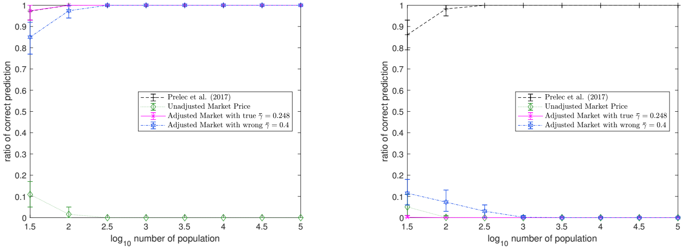


Figure 4: **The prediction performance when Assumption A7 may be invalid**. Note that $D_{KL}(f(\cdot|0), f(\cdot|1)) = 0.411 > \log \frac{7}{5}$ on the left, which means the bias in prior is not too large and two answers are still distinguishable, so our estimator is still consistent. While $D_{KL}(f(\cdot|0), f(\cdot|1)) = 0.082 < \log \frac{7}{5}$ on the right and our estimator fails to converge to the correct answer. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. The question is whether Philadelphia is the capital of Pennsylvania, and the correct answer is no, i.e. $\Omega_{obj} = 0$. The parameter is taken from the Philadelphia problem in Prelec et al. (2017). $\pi = (\frac{5}{12}, \frac{7}{12})$ and the objective likelihood $f(1|1) = \frac{20}{21}$, $f(1|0) = \frac{2}{3}$ for the left panel and $f(1|1) = \frac{5}{6}$, $f(1|0) = \frac{2}{3}$ for the right panel. Agents are assumed to have CRRA utility with $\gamma_i \sim Unif(0.1, 0.5)$, and initial wealth $w_i \sim Unif(0, 10)$. In the simulation, $M = 100$ times and $B = 100$ times. Our estimator is not consistent if Assumption A7 is invalid

Content

1 Introduction

2 Review of Bayesian truth serum

3 Design of a prediction market

4 Main results
- Hybrid Market-Survey Estimator
- Adjusted Market Estimator

5 Numerical results
- Hybrid estimator
- Adjusted market estimator

6 Conclusion

Conclusion

- We complement the survey based estimator in Prelec et al. (2017) by giving two estimators, a hybrid estimator based on market information and one survey question, and an adjusted market estimator.
- All these estimators are consistent under different sets of conditions, thus giving people more flexibility to choose which estimators to use.

Thanks for listening

Reference

Aurélien Baillon. Bayesian markets to elicit private information. *Proceedings of the National Academy of Sciences*, 114(30): 7958–7962, 2017.

Marco Ottaviani and Peter Norman Sørensen. Price reaction to information with heterogeneous beliefs and wealth effects: Underreaction, momentum, and reversal. *American Economic Review*, 105(1):1–34, 2015.

Dražen Prelec. A bayesian truth serum for subjective data. *Science*, 306(5695):462–466, 2004.

Dražen Prelec, H Sebastian Seung, and John McCoy. A solution to the single-question crowd wisdom problem. *Nature*, 541(7638): 532, 2017.

Justin Wolfers and Eric Zitzewitz. Interpreting prediction market prices as probabilities. Technical report, National Bureau of Economic Research, 2006.

Extensions and related issues

- Multiple-outcome events, risk averse agents
  - Consistency of hybrid estimator is unknown yet.
  - Adjusted market estimator works.
- Risk neutral agents
  - Hybrid estimator works in both binary and multiple case.
  - Adjusted market estimator fails.
- Robustness about the choice of risk aversion parameter
  - Hybrid estimator does not rely on the risk aversion parameter.
  - Adjusted market estimator is still consistent if the risk aversion parameter lies close to the true value.
- Incentivized design
  - The survey question can be incentivized to be close to the true prediction, if agent is close to risk neutral.
  - Strategies in prediction markets are automatically incentivized.

Theorem 2

In the multi-outcome case, our adjusted market estimator is given by

$$\hat{\Omega}_{obj,market} = \arg \max_{0 \le k \le K} \{\log \bar{p}^k + \frac{\bar{\gamma}}{N} \sum_{i=1}^{N} z_i^k\}, \tag{11}$$

where $z_i^k = x_i^k$ *for CARA utility and* $z_i^k = \log x_i^k$ *for CRRA utility. Under Assumptions A5, A6, and A7 and Regularity Conditions 1 and 2,* $\hat{\Omega}_{obj,Market}$ *is consistent, i.e.* $\hat{\Omega}_{obj,market} \to \Omega_{obj}$*,* $\mathbb{P}_{obj}$*-almost surely.*

Adjusted market estimator vs BTS: multiple-outcome case

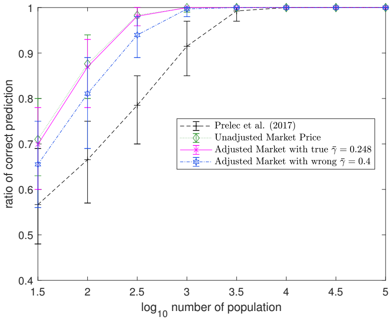


Figure 5: **The prediction performance when all assumptions are satisfied in multiple outcome case**. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. With our notation, the prior $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, the objective likelihood $f(\cdot|0) = (0.55, 0.1, 0.35)$, $f(\cdot|1) = (0.23, 0.75, 0.02)$, $f(\cdot|2) = (0.41, 0.13, 0.46)$, and the posterior probability $P_{0\cdot} = (0.4622, 0.1933, 0.3445)$, $P_{1\cdot} = (0.1020, 0.7653, 0.1327)$, $P_{2\cdot} = (0.4217, 0.0241, 0.5542)$. The agents are assumed to have CRRA utility function with risk aversion coefficient $\gamma_i \sim Unif(0.1, 0.5)$, and initial wealth $w_i \sim Unif(0, 10)$. The true state is $\Omega_{obj} = 0$. In the simulation, $M = 100$ and $B = 100$. Both estimators are consistent if all assumptions are valid.

Adjusted market estimator vs BTS: multiple-outcome case

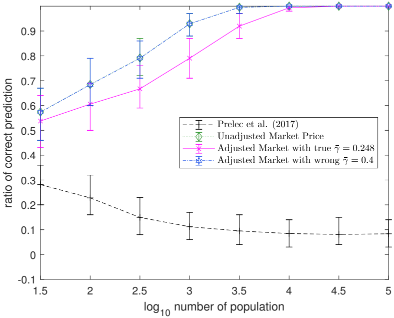


Figure 5: **The prediction performance when Assumption A1 is invalid**. In this case, assumption of common prior is invalid. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. With our notation, the prior π_i is symmetrically distributed around $\frac{1}{3}$, the objective likelihood is $f(\cdot|0) = (0.4993, 0.4645, 0.0361)$, $f(\cdot|1) = (0.4371, 0.5262, 0.0366)$, $f(\cdot|2) = (0.4305, 0.1710, 0.3984)$, and the posterior probability is $P_{0\cdot} = (0.3653, 0.3198, 0.3149)$, $P_{1\cdot} = (0.3998, 0.4530, 0.1472)$, $P_{2\cdot} = (0.0766, 0.0777, 0.8457)$. The agents are assumed to have CRRA utility function with risk aversion coefficient $\gamma_i \sim Unif(0.1, 0.5)$, and initial wealth $w_i \sim Unif(0, 10)$. The true state is $\Omega_{obj} = 0$. In the simulation, $M = 100$ and $B = 100$. The survey based BTS fails to lead to true state in this case, even if the population is very large. However, our adjusted market BTS estimator is still consistent.

Adjusted market estimator vs BTS: multiple-outcome case

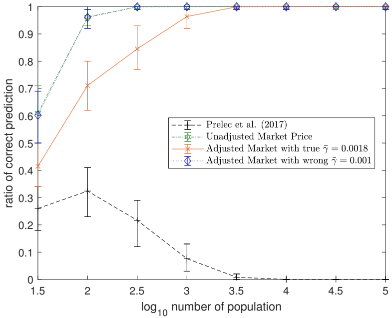


Figure 5: **The prediction performance when Assumption A3 is invalid**. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. With our notation, the prior $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, the objective likelihood $f(\cdot|0) = (0.4993, 0.4645, 0.0361)$, $f(\cdot|1) = (0.4371, 0.5262, 0.0366)$, $f(\cdot|2) = (0.4305, 0.1710, 0.3984)$, and the posterior probability $P_{0\cdot} = (0.3653, 0.3198, 0.3149)$, $P_{1\cdot} = (0.3998, 0.4530, 0.1472)$, $P_{2\cdot} = (0.0766, 0.0777, 0.8457)$. The agents are assumed to have CRRA utility function with risk aversion coefficient $\gamma_i \sim Unif(0.1, 0.5)$, and initial wealth $w_i \sim Unif(0, 10)$. The true state is $\Omega_{obj} = 0$. In the simulation, $M = 100$ and $B = 100$. In this case, Assumption A3 in Prelec et al. (2017) is invalid and BTS appears inconsistent while our estimator still converges to the correct answer.

Adjusted market estimator vs BTS: multiple-outcome case

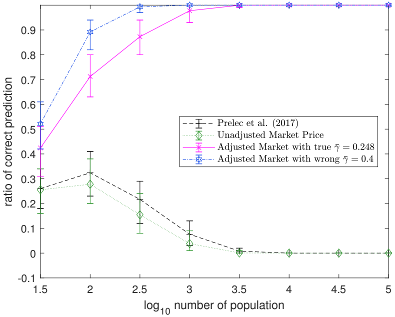


Figure 5: **The prediction performance when Assumption A4 is invalid**. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. With our notation, the prior $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, the objective likelihood $f(\cdot|0) = (0.2565, 0.1937, 0.5498)$, $f(\cdot|1) = (0.2671, 0.2827, 0.4502)$, $f(\cdot|2) = (0.2019, 0.2127, 0.5854)$, and the posterior probability $P_{0\cdot} = (0.3535, 0.3682, 0.2783)$, $P_{1\cdot} = (0.2811, 0.4102, 0.3087)$, $P_{2\cdot} = (0.3468, 0.2840, 0.3692)$. The agents are assumed to have CRRA utility function with risk aversion coefficient $\gamma_i \sim Unif(0.1, 0.5)$, and initial wealth $w_i \sim Unif(0, 10)$. The true state is $\Omega_{obj} = 0$. In the simulation, $M = 100$ and $B = 100$. In this case, Assumption A4 in Prelec et al. (2017) is invalid and BTS appears inconsistent while our estimator still converges to the correct answer.

Adjusted market estimator vs BTS: multiple-outcome case

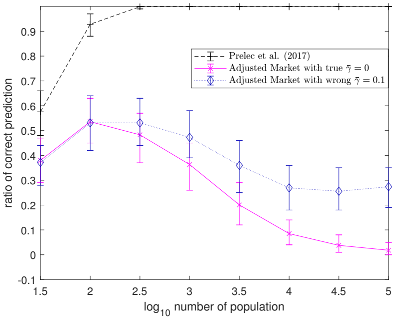


Figure 5: **The prediction performance when Assumption A5 is invalid**. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. With our notation, the prior $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, the objective likelihood $f(\cdot|0) = (0.55, 0.1, 0.35)$, $f(\cdot|1) = (0.23, 0.75, 0.02)$, $f(\cdot|2) = (0.41, 0.13, 0.46)$, and the posterior probability $P_{0\cdot} = (0.4622, 0.1933, 0.3445)$, $P_{1\cdot} = (0.1020, 0.7653, 0.1327)$, $P_{2\cdot} = (0.4217, 0.0241, 0.5542)$. The agents are assumed to have CRRA utility function with risk aversion coefficient $\gamma_i \sim Exp(1)$ and initial wealth $w_i \sim Unif(0, 10)$. The true state is $\Omega_{obj} = 0$. In the simulation, $M = 100$ and $B = 100$. Assumption A5 is invalid since $\mathbb{E}_{obj}[\frac{1}{\gamma_i}] = \infty$. Our estimator appears to be inconsistent.

Adjusted market estimator vs BTS: multiple-outcome case

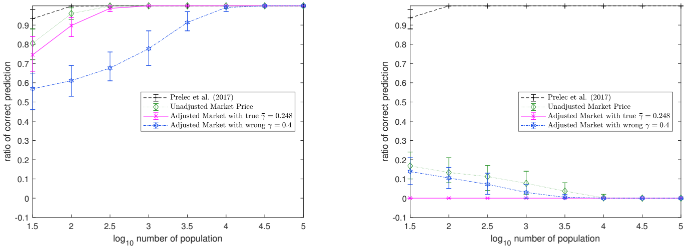


Figure 6: **The prediction performance when the Assumption A7 is invalid**. Note that $D_{KL}(f(\cdot|2), f(\cdot|1)) = 1.451 > \log 2$ on the left, which means the bias in prior is not too large and two answers are still distinguishable; our estimator appears to be consistent. While $D_{KL}(f(\cdot|0), f(\cdot|1)) = 0.298 < \log 2$ on the right, and our estimator fails to converge to the correct answer. The upper and lower bar in the figure stands for the 95% confidence interval of the ratio of getting correct predictions in 100 tests. With our notation, the prior $\pi = (0.25, 0.5, 0.25)$ and the objective likelihood $f(\cdot|0) = (0.55, 0.1, 0.35)$, $f(\cdot|1) = (0.23, 0.75, 0.02)$, $f(\cdot|2) = (0.41, 0.13, 0.46)$ in the left panel while $f(\cdot|0) = (0.61, 0.09, 0.3)$, $f(\cdot|1) = (0.3, 0.4, 0.3)$, $f(\cdot|2) = (0.1, 0.25, 0.65)$ in the right panel. The agents are assumed to have CRRA utility function with risk aversion coefficient $\gamma_i \sim Unif(0.1, 0.5)$, and initial wealth $w_i \sim Unif(0, 10)$. The true state is $\Omega_{obj} = 2$ in the left panel and $\Omega_{obj} = 0$ in the right panel. In the simulation, $M = 100$ and $B = 100$. In this case, the prior distribution is no longer symmetric. Our adjusted market estimator is not always consistent, depending on whether Assumption A7 holds or not.

Proposition 3

(a) *Given the market price* $\bar{p} \in \mathbb{R}^{K+1}$*, agent* i*'s optimal strategy is*

$$x_i^k = \frac{w_i}{\bar{p}^k} \mathbf{1}_{\{k = \arg\max_{0 \le j \le K} \frac{p_i^j}{\bar{p}^j}\}}. \tag{12}$$

That is, the agent will invest all of his money in one asset that has the lowest price relative to his subjective probability.

(b) *Under Assumptions A1, A2 and A3, the equilibrium exists if and only if* $k = \arg\max_{0 \le j \le K} \frac{P_{kj}}{\bar{p}^j}$ *for all* $0 \le k \le K$*, where* $P_{kj} = \mathbb{P}(\Omega = j | S_i = k)$ *and equilibrium price* $\bar{p}$ *is given by*

$$\bar{p}^k = \frac{1}{\sum_{i=1}^N w_i} \sum_{i:S_i = k} w_i. \tag{13}$$

Theorem 3

Consider the estimator defined by

$$\hat{\Omega}_{neutral} = \arg\max_{0 \leq k \leq K} \sum_{i=1}^{N} \hat{v}_i^k \sum_{l=0}^{K} \frac{\hat{m}_{kl}}{\hat{m}_{lk}}, \tag{14}$$

where $\hat{v}_i^k = \mathbf{1}_{\{x_i^k = \max_{0 \leq j \leq K} x_i^j\}}$ *and* $\hat{m}_{kl} = \frac{1}{\sharp\{i : \hat{v}_i^k = 1\}} \sum_{i : \hat{v}_i^k = 1} \xi_i^l$. *Suppose the equilibrium market price exists. Then under Assumptions A1, A2, A3 and Regularity Condition 1, the above estimator is consistent, i.e.* $\hat{\Omega}_{neutral} \to \Omega_{obj}$ *a.s.* $\mathbb{P}_{obj}$.

Proposition 4

Suppose Assumptions A5, A6, A7 hold true. If there exists $m > 0$, $M > 0$, *such that* $-m \leq \liminf_{N\to\infty} \log \frac{\bar{p}^{\Omega_{obj}}}{\bar{p}^j}$, $\limsup_{N\to\infty} \log \frac{\bar{p}^{\Omega_{obj}}}{\bar{p}^j} \leq M$, $\mathbb{P}_{obj}$*-a.s. for any* j, *then our estimator* (10) *is still consistent with a wrong average risk aversion* $\tilde{\gamma}$ *that satisfies*

$$\frac{\bar{\gamma}}{1+\frac{d}{m}} < \tilde{\gamma} < \frac{\bar{\gamma}}{(1-\frac{d}{M})^+}, \tag{15}$$

where $d = \min_{j\neq\Omega_{obj}}\{D_{KL}(f_{obj}(\cdot|\Omega_{obj}), g(\cdot|j)) + \log \frac{\bar{\pi}^{\Omega_{obj}}}{\bar{\pi}^j}\} > 0$.

Proposition 5

Suppose the payoff of answering the survey question to agent i is $r(\xi^1, n_{-i}) = 1 + \epsilon - (\xi^1 - n_{-i})^2$, where $n_{-i} = \frac{1}{N-1}\sum_{j \neq i} \mathbf{1}_{\{x_j > y_j\}}$ for agent i. If $0 < \gamma_i \ll 1$, then $\xi_i^1 = \bar{n}_{-i} + c\gamma_i + O(\gamma_i^2)$, where $\bar{n}_{-i} = \mathbb{P}(S_j = 1 | S_i)$. In particular, ϵ can be chosen such that $|c| < 0.1$. Moreover, if $\gamma_i = 0$, then $\xi_i^1 = \bar{n}_{-i}$

- a quadratic payoff function relates to conditional expectations;
- answers are still close to one's truthful predictions if one is risk averse;
- agents will not deviate from the strategy (x_i, y_i, ξ_i^k) in markets and survey questions.