Efficiency of the Experimental Prediction Market: Public Information, Belief Evolution, and Personality Traits

Bin-Tzong Chie[1] and Chih-Hwa Yang[2]

Abstract

This paper examines the ability of markets to aggregate information so that the price generated from the market contains the best estimate of all the available information. The paper investigates how individuals "update" their initial beliefs from their public and private information in light of market prices. In particular, the paper looks at individuals' weighting of public information versus private information. Also, the volume of information in the market via an increased number of traders with private information has a positive impact on the quality of the market price. Lastly, the personality traits of the traders seem to provide some positive impact if the traders are diverse in terms of the proportion of "efficient and organized" traders in the market.

---

[1] Department of Industrial Economics, Tamkang University, Taiwan.
[2] Department of Industrial Economics, Tamkang University, Taiwan.

1. Introduction

Economists have tried to identify possible causes of market efficiency. The prediction market applies a trading mechanism to disclose and turn fragmented information into a collective market price. George Neumann, Robert Forsythe and Forrest Nelson started to utilize market mechanism to predict the election outcomes in 1988. They established the Iowa political stock market that allowed faculty, staff and students to trade at the University of Iowa. What was being traded at that time were the futures contracts of voting shares for George H. W. Bush, Michael Dukakis, and the other presidential candidates. The results had shown that the prediction market was rather accurate, outperforming concurrent polls (Hahn and Tetlock 2006). In 1992, the Iowa political stock market was renamed the Iowa Electronic Market (IEM) and traded a variety of objects with the mechanism of futures market. As the networking infrastructure became advanced, the IEM participants could come from all over the world.

The efficiency of the prediction market has been supported empirically and theoretically by Wolfers and Zitzewitz (2004) and Gjerstad (2005). Even a small scale market can be efficient (Christiansen, 2007). Researchers have endeavored to unravel information efficiency embedded in the market. For example, Forsythe et al. (1992) and Forsythe, Rietz and Ross (1999) successfully predicted the outcome of the US presidential election with the political stock market.

However, it remains questionable whether the market can always aggregate information efficiently. Before the result of the U.S. presidential election of 2016, the IEM still predicted that Hillary Clinton will beat Donald Trump. Haruvy, Lahav and Noussair (2007) found that market individuals' beliefs about prices are informative for researchers to infer the movements of future market price. While remaining strong beliefs, the IEM should nevertheless offer a false prediction. When the traders had gradually realized the electoral outcome Hillary's contract began to crash and Trump's contract suddenly became very hot.

On the other hand, Ottaviani and Sørensen (2007) suggested that under certain circumstances it might be difficult for the market to achieve efficiency, such as no trade theory, herd behavior and manipulation. In this paper, our goal is to apply an experimental prediction market to revisit market efficiency and explore the influence of individual beliefs and personality traits under an experimental market. Through market mechanism, many researchers indicate that the advantage of markets. Hanson (2006) has introduced many predictive phenomena that have long existed in markets. For example, orange-juice futures can help the National Weather Service improve weather forecasts (Roll, 1984); horse race betting beats experts' forecasts (Figlewski, 1979); stock markets reacted earlier than the NASA investigation panel in pointing to the company responsible for the Challenger accident (Maloney and Mulherin, 2003). These findings support the functionality of information aggregation of markets and the importance of information disclosure from market participants.

The rest of the paper is organized as follows. Section 2 reviews the literature of

prediction markets. Sections 3 and 4 detail our experimental design, data analysis. Section 5 presents the results and a discussion, and is followed by the conclusion in Section 6.

2. Literature Review of Prediction Markets

Prediction markets apply the market mechanism to achieve the information aggregation. With participants motivated by profitable opportunity, the prediction market not only provides up-to-date information, but also saves tremendous cost and time compared to polls. In other words, by properly aligning the incentives of participants, the prediction market can reveal true preferences at the aggregate level economically. Berg, Nelson, and Rietz (2008) demonstrated that the IEM outperformed polls for 76 percent of the elections between 1988 and 2004. Snowberg, Wolfers, and Zitzewitz (2007) studied a longer period of election samples and found a stable pattern of electoral outcomes affecting financial markets. These findings suggest that markets would try to account for valuable information by all means. The information advantage has been studied through the data of the IEM (Berg et al., 2008). In particular, Berg et al. (2008) found that some traders might have more information than others, being defined as "marginal traders". As the marginal traders search for profitable opportunities, they improve the market efficiency.

Many studies have tried to identify possible factors of market efficiency. Figlewski (1978) showed that it is mathematically evident that the more homogeneously distributed information is, the more efficient a prediction market can be. Oliven and Rietz (2004) analyzed the questionnaire data of the IEM traders and found that despite of participants' anomalous and irrational behavior, the prediction market can be as efficient. These findings corroborate the Hayek hypothesis. According to Hayek (1945), although bearing bias, the market is able to aggregate individual opinion and form a "price" that not only conveys information but also directs resource allocation.

On the contrary, there are findings for market inefficiency (Rosenberg, Reid and Lanstein, 1985; Hazen, 1987). One of the noticeable facts on market inefficiency is the systematic pricing bias, which is called the favorite-longshot bias (FLB). This bias is attributed to the traders' tendency in overestimating low probability events or underestimating high probability events (Griffith, 1949). Isaacs (1953) demonstrated that the bias resulted from the profit maximization behavior of the monopoly informed trader. Weitzman (1965) discussed the willingness of risk-seeking traders to accept a lower expected payoff, causing over investment in low probability events. Heterogeneous beliefs among traders might also contribute to the bias (Ali, 1977).

Ottaviani and Sørensen (2010) developed a series of testable features, including the quality of private information, the number of traders, the number of outcomes, common unpredictable errors, the level of participation, and the prior distribution of beliefs. These have provided us with clues for constructing the experimental

market.

There are studies that investigate manipulators in experimental markets. Deck et al. (2013) showed that market participants can be distinguished by the level of information possessed. Informed traders (the manipulators) have complete information and often act in a disguised manner in an attempt to dictate the market. Ordinary traders are those who own partial information and adjust views based on observed market prices. In addition to manipulators and ordinary traders, there are forecasters who have no information at all and are often beguiled by the behavior of manipulators. Deck et al. (2013) have shown that the interaction between manipulators and forecasters results in market failure.

As shown in the above literature discussion, there exists no unanimous agreement on market efficiency. In this paper, we attempt to control market variables by setting up an experimental prediction market in accordance with Deck, Lin and Porter (2013). In our small-scale experimental prediction markets, we test market efficiency by altering factors such as the distribution of private information, initial beliefs, public information, the number of traders, the evolution of beliefs, and the heterogeneity of personality traits. This study will not focus on the effects of manipulators. Our experimental market consists of ordinary traders only. Through the double auction mechanism, we intend to observe market performance without fully informed traders so as to investigate other possible endogenous causes for efficiency or inefficiency.

3. Experimental Design

This section will first introduce our subjects, experimental design followed by a description of the distribution of information and beliefs.[3]

3.1   Subjects

We conducted 15 experimental sessions and had totally recruited 183 subjects.[4] The subjects were college or graduate students recruited from universities in Taiwan. Each subject only attended one of our experimental sessions. Repeat participants were not allowed. The number of subjects (m) per experiment session ranged from 5 to 19. Each session lasted for 2.5 hours, including 1 hour of introduction, 1 hour of trading, and 0.5 hours of personality testing. The average pay rate for our experiment is about NT$160 per hour. The total actual reward depends on performance, ranging from NT$260 to NT$500, with the average being NT$400 ($2.5 \times 160$).

---

[3] The experimental data and subject instruction can be obtained from the authors upon request.

[4] The experiments were constructed using z-trees (Fischbacher 2007) and conducted in the Experimental Economics Lab at National Chengchi University. Further information about the institution can be found at http://futures.nccu.edu.tw/.

3.2   Information Distribution Design

Through the pre-experiment introduction, every subject was made aware of the fact that information is valuable. In our experiment, the information distribution of each session was represented by a set of N balls in a black box, which were k yellow balls and $(N - k)$ white balls. Each subject was asked to draw n balls with replacement from the black box and keep the information privately. The drew n balls were x yellow balls and $(n - x)$ white balls. After all subjects' drawing, the experimenter then drew a ball from the black box and announced the color to all subjects. The ball was represented by y, which was 1 if the ball was yellow, otherwise, y was 0. After that, subjects were instructed to predict the percentage of yellow balls. For each participant, according to each $p = \frac{k}{N}$, the probability of observed x (private information) plus y (public information) yellow balls can be represented by the following probability mass function:

$$f(x, y; n, p) = \Pr(X = x + y) = \binom{n}{x} p^{x+y}(1 - p)^{n+1-x-y},$$

$$x = \{0, 1, \dots, n\} \; ; \; y = \{0,1\}. \tag{1}$$

In our experiment, $N = 10$, $n = 2$, $k = 1, 2, \dots, 9$.

3.3   Beliefs of Subjects

After observing the private and public information, all subjects were requested to form and report their beliefs. For example, if a subject observed two yellow balls and the public information was also a yellow one, he or she might think that over 80 percent of the balls in the black box were yellow. If the market price fell below 80, this participant would go long. For our analysis, each subject had to answer the two questions below:

1. How many yellow balls did you observe, including your draw and the public announcement?
2. What do you believe is the percentage of yellow balls contained in the black box?

Trading was launched after all subjects completed above two questions.

According to the probability mass function (1), suppose the prior probabilities of $k = 1, 2, 3, \dots, 9$ balls were uniform.
We can infer subjects' posterior beliefs predicted by the following Bayesian updating.

$$\mathrm{P_r}(k = i | X = x + y) = \frac{\mathrm{P_r}(X|k=i)\mathrm{P_r}(k=i)}{\mathrm{P_r}(X)} = \frac{\mathrm{P_r}(X|k=i)\mathrm{P_r}(k=i)}{\sum_{j=1}^{9}\mathrm{P_r}(X|k=j)\mathrm{P_r}(k=j)}, \tag{2}$$

where $i = 1, 2, \dots, 9$.

Then the expected posterior beliefs for the number of yellow balls can be inferred by $E(k|X = x + y) = \sum_{j=1}^{9}(k = j|X = x + y)\,k$. The Bayesian percentage of yellow balls contained will be $10 \times E(k|X = x + y)$.

3.4   Trading

Each session had two treatments: one was based on the ratio of yellow balls (i.e., the *ratio treatment*) and the other on whether yellow balls outnumbered white ones (i.e., the *winner-take-all treatment*). The experimenter should announce which treatment he will begin with. Each treatment lasted for four rounds and each round took three minutes. Based on the true value K ($= 10 \times k$), the market price was defined, lying between 0 (i.e., no yellow ball) and 100 (i.e., all yellow balls).

Each round, traders will be lending 1,000 tokens from the experimenter. Traders are allowed to buy or sell with these tokens. At the end of the market round, the 1,000 tokens will be returned to the experimenter. In this study, the trading mechanism of the futures market is applied. Subjects' contract holding position can be positive or negative. Positive means net buying, otherwise means net selling. During the trading period, the subjects' token holding is just temporary, for that all contract holding will also be settled with token at the end of each round. The goal of each subject is try to use all available information to trade. The final payoff is a linear transformation of the tokens subjects earn.

A standard continues double auction trading mechanism is adopted in this experimental market. For both treatments, the market reference price started at 50, representing maximum uncertainty of the percentage of yellow balls. Under the double auction mechanism, subjects could place their bid or ask prices by integer ranging from 0 to 100, or accept what was offered by other traders. After subjects placed their bid or ask prices, there were several bid and ask prices; they were maintained in the bid and ask queues, ordered first by offer price and then by time of placement.

The maximum possible loss will be deducted first from the token account for each traded contract. For example, If a contract price of a transaction is 60, the buyer pays 60, and the seller pays 40 as a deposit. If the final settlement price is 80, the buyer will receive 80 and the seller will only receive 20 tokens back. If the subject does not have enough tokens, no further orders can be placed.

At the end of the trading, subject i's post-transaction belief (b_i^p), which would be associated with additional reward or punishment, was elicited. The incentive was equal to $50 - |b_i^p - K|$ in token value. This settlement process took place after all subjects declared their post-transaction beliefs. Each round of the market clearing was a zero-sum game.

3.5   Settlement

The experimental prediction market applied a cash settlement. For the settlement price (S) and holding position (h_i), trader i was to receive ($S \times h_i$) in token value. Here, S was dependent on the type of the two treatments applied. For the *winner-*

take-all (WTA) *treatment*, the settlement price S was determined as follows:

$$S = \begin{cases} 100, & \text{if } K > 50; \\ 50, & \text{if } K = 50; \\ 0, & \text{if } K < 50. \end{cases}$$

(3)

For the *ratio treatment*, the settlement price S was defined as $S = K$.

4. Experimental Data Analysis

A total of 15 sessions were conducted; each session had 8 rounds. In this study, the characteristic features of each round (r) was calculated as a data record. Therefore, the sample size of the experimental data was 120.

We tried to measure market efficiency by means of the following procedures. The purpose of this experiment is to compare how the market prices compare to predicted market prices, presumably based on the objective probabilities of the number of yellow balls available.

We first define the objective probabilities that prices should be compared to. For example, if there are 7 subjects and they observe 9 yellow balls and 6 white balls, what is the probability that there were $k = 1, 2, 3, …, 9$ yellow balls in the black box? The answer to that question is determined by Bayes' theorem. Then the probability that draws are made from a black box with k yellow balls given the sample

$$\mathrm{P_r}(k = i | p) = \frac{\mathrm{P_r}(p|k=i)\mathrm{P_r}(k=i)}{\mathrm{P_r}(p)} = \frac{\mathrm{P_r}(p|k=i)\mathrm{P_r}(k=i)}{\sum_{j=1}^{9}\mathrm{P_r}(p|k=j)\mathrm{P_r}(k=j)},$$

(4)

where $i = 1, 2, …, 9$ and $p = \frac{k}{10}$.

For each $k = j$, where $j = 1, 2, …, 9$, the probability of each of these samples is binomial distribution $X \sim b(n, p)$ with $n = 15$, $p = j / 10$, and d equal to the number of yellow balls drawn in the sample. The probability of drawing d yellow balls in a sample of size n conditional on k yellow balls in the black box is $f(d, n, p) = \mathrm{P_r}(X = d) = \binom{n}{d}(p)^d(1 - p)^{(n-d)}$.

In the example we took this as $d = 9$. Suppose the prior probabilities of $k = 1, 2, 3, …, 9$ balls were uniform. All priors are 1/9. So in the example, the probabilities are (0.0000, 0.0011, 0.0185, 0.0979, 0.2444, 0.3306, 0.2356, 0.0688, 0.0031). In the example with group size seven the expected value of the number of balls in the black box is 0.5882 when 9 of 15 balls drawn are yellow, and that is true regardless of the number of yellow balls in the black box. Suppose there are five yellow balls in the black box but 9 of the 15 draws are yellow. Then the measure that the we use as the *fundamental value* (*FV*) is what the subjects could reasonably infer given the information available to them.

Then, we used closing price (p_c) of each round as a proxy of market performance.[5] We define the *fundamental-based pricing error* as p_c minus the respective fundamental value. Figure 1 shows the *pricing error* ($p_c - FV$) (y-axis) versus each *true value* (K) (x-axis) for both treatments. The vertical bars mark the range of the mean plus and minus one standard deviation.

4.1   Market Efficiency

From Figure 1, we observe diverging slopes between the two treatments. The following regression will be applied to test whether the coefficients of treatment dummy and designed true value (K) are significant or not.

$$Error = \beta_0 + \beta_1 D_t + \beta_2 K + \beta_3 D_t K + \varepsilon ,\qquad(5)$$

where ε is assumed to be an i.i.d. random variable that is normally distributed and D_t is the treatment dummy ($D_t = 1$ for the WTA treatment, and 0 for the ratio treatment).
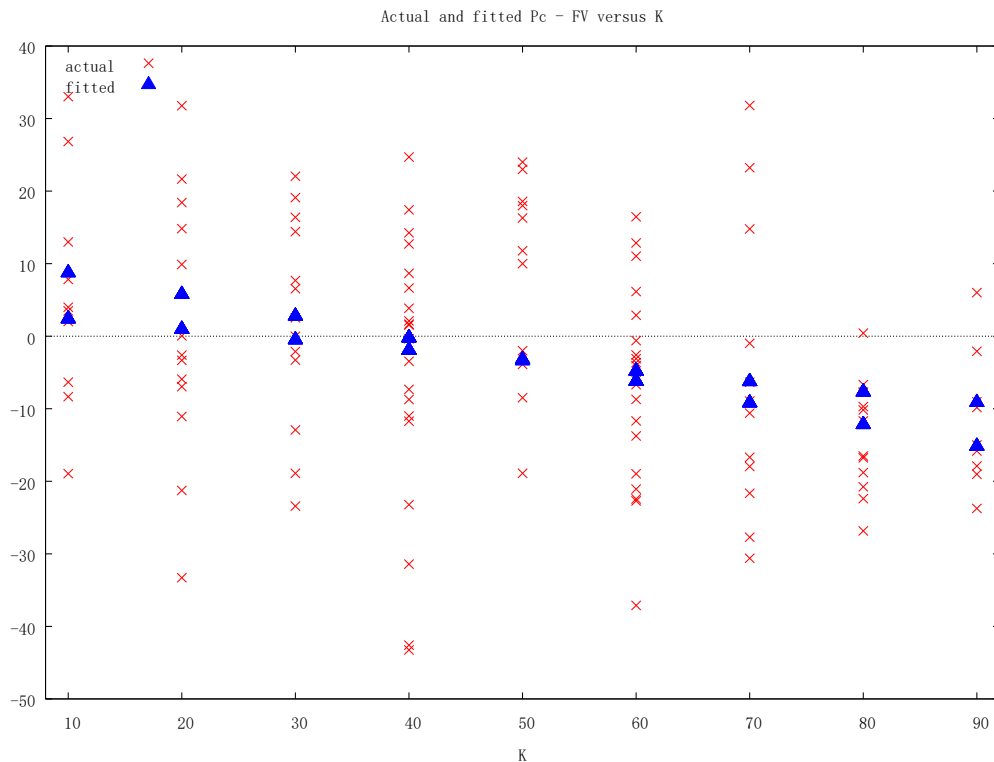

Figure 1: The pricing error (pc – FV) (y-axis) versus each true value (K) (x-axis) for both treatments

---

[5] In fact, Deck et al. (2013) justify using closing prices (in addition to mean prices) by stating, "previous studies have found that closing prices typically better reflect aggregated information."

Below is the regression result of our experimental data.

$$Error = \underset{(0.0263)}{11.751} - \underset{(0.2571)}{7.912}\, D_t - \underset{(0.0007)}{0.299}\, K + \underset{(0.2102)}{0.155}\, D_t K, \quad \bar{R}^2 = 0.09,$$

where the *p*-values are presented in the parentheses and $\bar{R}^2$ is the adjusted R^2. For the ratio treatment ($D_t = 0$), the constant term (11.751) shows that the market price tends to overprice the fundamental value when K is small. As K increases, *Error* will be alleviated until K reaches a critical value, which is around 39.3 (=11.751/0.299). When K is larger than 39.3, the market will tend to underprice the fundamental value. The WTA treatment also suffers, but non-significant with a milder constant term 3.839 (= 11.751 – 7.912) and slope -0.144 (= -0.299 + 0.155). Although this pattern may compliant with the FLB, the treatment and interaction dummies are not significant. We simply apply regression (6) to test whether the coefficient of designed true value is significant or not.

$$Error = \beta_0 + \beta_1 K + \varepsilon, \tag{6}$$

The result is presented below

$$Error = \underset{(0.1029)}{7.589} - \underset{(0.0019)}{0.219}\, K, \quad \bar{R}^2 = 0.09.$$

Through the low $\bar{R}^2$, coefficient of determination, shows less than 10% of the variance in the error that is explained by the designed true value. However, the perception of participants may also be subject to their own beliefs.

For the error measurement of the aforementioned influence from perception, we define the following variables:

- Mean initial belief: $\bar{b} = \sum_{i=1}^{m} \dfrac{b_i}{m}$, where b_i is the initial belief of subject *i*; and

- Mean post-transaction belief: $\bar{b}^p = \sum_{i=1}^{m} \dfrac{b_i^p}{m}$, where b_i^p is the belief reported by subject *i* at the end of transaction.

The null hypothesis we proposed here is that the pricing error is a pure white noise, which is

$$H_0: \quad \beta_0 = \beta_1 = 0$$

Table 1: Market efficient test

Error Measurement (dependent variable)		$\hat{\beta}_0$	$\hat{\beta}_1$	$\overline{R}^2$
(a)	$p_c - FV$	7.589 (0.1029)	-0.219*** (0.0019)	0.09
(b)	$p_c - \bar{b}$	3.221 (0.4225)	-0.1261* (0.0488)	0.05
(c)	$p_c - \bar{b}^p$	2.414 (0.3787)	-0.092* (0.0678)	0.05
(d)	$\bar{b} - FV$	4.368* (0.0978)	-0.093** (0.0452)	0.05
(e)	$\bar{b}^p - FV$	5.175* (0.0936)	-0.127** (0.0105)	0.06
(f)	$\bar{b}^p - \bar{b}$	0.807 (0.6274)	-0.025 (0.1475)	0.01

Note: *, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively. *p*-values in parentheses.

The test results are shown in Table 1. The error measurements of (a) to (c) failed to reject $\beta_0 = 0$. However, if we look at β_1, the slope of the regression line, it can also be interpreted as a proxy of the degree of market efficiency. The smaller $\widehat{\beta_1}$ is in absolute value terms, the more efficient the market is. According to Table 1, the market price is more efficient in reflecting the mean post-transaction belief ($\bar{b}^p$) than the fundamental value (*FV*).

If we further take look at (d) and (e), we will have $\bar{b} - FV$ and $\bar{b}^p - FV$. All coefficients are significantly different from zero. This means that belief could be the reason to cause the market price deviate from its fundamental value. This might be the case where the majority of subjects did not take the sampled observations as their beliefs. They tended to attach too much weight to the probability of the lower observations and too less weight to the higher ones. This phenomenon can be captured by the probability weighting function in Tversky and Kahneman's (1992) study and followed by Camerer and Ho (1994), Tversky and Fox (1995), Wu and Gonzalez (1996), Abdellaoui (2000), Abdellaoui, Vossmann and Weber (2005), and Stott (2006).

However, the probability weighting function may not be the only reason for the market efficiency distortion. The bargaining process might also play a role (Knetsch, Tang and Thaler 2001; Loomes, Starmer and Sugden 2003; Tufano 2010). Although (f) $\bar{b}^p - \bar{b}$ only posts a small deviation, this might suggest that the mean market price does not perfectly converge to either the mean initial belief or the mean post-transaction belief. The subjects' bid or offer might not always be consistent with their own beliefs, which are affected by the transaction price from time to time. Some are more open to making aggressive offers, while some appear hesitant to take the risk. It is observed that heterogeneous subjects can generate different

pricing behaviors.[6] If all subjects place orders in accordance with their beliefs, the market price will be fairly close to the mean belief. In other words, when all traders are honest, the market price will be representative of the joint true beliefs.

Nevertheless, we have observed that the market price did not converge to the subjects' beliefs or observations. From our experimental design, the degree of honesty of individual subjects might be further related to the results of the personality traits test. In the following subsection, we will introduce a general model while considering the personality traits for further hypothesis testing.

4.2   Panel Regression Model

In the empirical model, the dependent variables used in the regressions may be associated with two market pricing errors: the first one is the deviation from the fundamental value (*FV*), ($p_c - FV$) (fundamental-based error); and the second one is the deviation from the mean post-transaction belief ($\bar{b}^p$), ($p_c - \bar{b}^p$) (belief-based error). The explanatory variables for both of our regression models are the treatment dummy (D^T), public information dummy (D^p), number of traders (*m*), gender ratio (*g*), and the limit order ratio (*l*) of each round.[7] The gender ratio *g* is defined by the portion of male subjects. For any individual trader, the limit order ratio tracks the number of limit orders over the total orders placed by the trader. The basic pricing error model using panel estimation for the two error measurements is extended as shown in the following regression:

$$Error = \alpha_t + \beta_1 D_{it}^T + \beta_2 K_{it} + \beta_3 D_{it}^T K_{it} + \beta_4 D_{it}^p + \beta_5 m_i + \beta_6 g_i + \beta_7 l_{it} + \varepsilon_{it}, \qquad (7)$$

where the subscript *it* denotes session *i* of round *t*, and α_t is the round-specific factor, which is a fixed value representing the characteristics of a specific round. The degree of market efficiency can be tested through α_t. The *experience level* of traders may be improved through repeated rounds (*t*). According to the above setting, we will analyze the round (time) fixed effects regression model.[8]

4.3   Heterogeneity of Personality Traits

Personality traits may provide valuable information regarding the individual-specific characteristics of subjects. In our experiment, we did not inflict controls on

---

[6] Several studies have discussed the behaviors of heterogeneous subjects in terms of culture (Oosterbeek, Sloof and Van De Kuilen, 2004; Ehmke, 2006), social preferences (Oechssler, 2012), gender (Dreber, von Essen and Ranehill, 2011; Croson and Gneezy, 2009) and personality (Visser and Roelofs, 2011; Ballinger et al., 2011).

[7] Forsythe et al. (1992) made a similar assumption.

[8] If we analyze the session fixed effects regression model, the effect of some variables which do not change with different rounds is unobserved. As such, no comparison is made between the results of the session fixed effects regression model and pooling model. Besides, the random effects regression model does not clearly account for the effect of increasing familiarity in the trading mechanism on market efficiency.

the traits of subjects participating in any experimental market; instead, we tried to identify possible characteristics of each session, with an attempt to understand the relationship between the heterogeneity of personality traits and market performance. The Five-Factor Model has been applied to our study. It characterizes people based on five dimensions measured by the NEO Personality Inventory in Costa and McCrae (1985), namely, neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness.

The Chinese version of the NEO Personality Inventory test has been made available by Dr. Julian Lai from the City University of Hong Kong.[9] The score range for each dimension is between 10 and 50. We calculated the standard deviation of each personality trait among participants in every session so that the difference between sessions can be captured. These heterogeneous measures of personality traits help distinguish the following opposite characteristics,

Neuroticism: *sensitive*/*nervous* as opposed to *secure*/*confident*
Extraversion: *outgoing*/*energetic* as opposed to *solitary*/*reserved*
Openness: *inventive*/*curious* as opposed to *consistent*/*cautious*
Conscientiousness: *efficient*/*organized* as opposed to *easy-going*/*careless*
Agreeableness: *friendly*/*compassionate* as opposed to *cold*/*unkind*

Table 2 provides summary statistics of the variables.

Table 2: Descriptive statistics

Variable	Min.	Max.	Mean	S.D.
Fundamental-based error ($p_c - FV$)	-43.25	33.02	-3.20	16.29
Belief-based error ($p_c - \bar{b}^p$)	-26.25	23.12	-2.09	9.18
Fundamental value (*FV*)	11.01	88.08	49.76	21.96
Number of traders (*m*)	5	19	12.20	4.17
Gender ratio (*g*)	0.31	0.88	0.48	0.14
Limit order ratio (*l*)	0.32	0.73	0.51	0.07
Public information dummy (D^p)	0	1	0.57	0.50
Mean initial belief ($\bar{b}$)	10	98	49.55	20.88
Mean post-transaction belief ($\bar{b}^p$)	8.80	98	48.66	20.48
Closing price (p_c)	0.00	100.0	46.56	21.25
S.D. of extraversion	1.47	7.37	5.29	1.34
S.D. of agreeableness	2.91	5.58	3.82	0.67
S.D. of conscientiousness	1.36	6.27	4.66	1.20
S.D. of neuroticism	2.94	8.59	6.19	1.43
S.D. of openness	3.74	6.97	5.25	0.93

---

[9] The original personality inventory test and evaluation can be requested from
http://www6.cityu.edu.hk/stfprofile/julian.lai.htm

4.4   **Hypotheses**

Based on the pattern of the pricing error, we found that systematic bias is prevalent in the experimental prediction market. Section 4.1 pointed out that error can be sourced from the mean belief and fundamental values. Empirical studies have proved that the number of traders, gender, and limit order ratio are deterministic in this regard. In addition to these variables, we analyzed the characteristics of *public information* and the *heterogeneity of personality traits* in both fundamental- and belief-based models. To be specific, we tried to test if the following hypotheses can withstand scrutiny.

H$_1$: Positive public information leads to market overpricing.
H$_2$: An increase in the number of traders can improve market efficiency.
H$_3$: An increase in the proportion of male participants can improve market
       efficiency.
H$_4$: An increase in the ratio of limit orders can improve market efficiency.
H$_5$: An increase in the heterogeneity of personality traits can improve market
       efficiency.

5. Results and Discussion

As can be seen in Tables 3 and 4, both error measurements were significantly influenced by true values (Ks). The signs and patterns of the coefficients were consistent with the results presented in Table 1. According to the estimation between the *pooled OLS* and the *fixed effects* regression in Table 3 and Table 4, we observed that there was very little difference between these two models, except for the variables omitted due to exact collinearity of the fixed effects model in Table 3. We will thus focus here on reporting the more informative results of the fixed effects regression. By comparing Model 1' with Model 2' and Model 3' with Model 4', we found that the coefficients of both error measurements were robust and free from the influence of the heterogeneity of personality traits. The same conclusion can be drawn from the error measurements (a) and (c) in Table 1. By comparing (a) with Model 1' and 2' and (c) with Model 3' and 4', most of the coefficients were significant except for the intercept. In the following, we will discuss the testing results of our earlier hypotheses.

Table 3: Regression results of fundamental-based error

	Pooled OLS		Fixed Effects	
	Model 1	**Model 2**	**Model 1'**	**Model 2'**
Intercept	16.37 (14.01)	36.05* (20.39)	18.07 (15.32)	18.07 (15.32)
D^T	-6.83 (6.10)	-7.89 (6.86)	-7.27 (6.89)	-7.27 (6.89)
K	-0.51*** (0.06)	-0.53*** (0.06)	-0.54*** (0.06)	-0.54*** (0.06)
D^T*K	0.10 (0.10)	0.12 (0.12)	0.11 (0.11)	0.11 (0.11)
D^p	19.66*** (2.54)	17.45*** (2.63)	17.49*** (2.92)	17.49*** (2.92)
m	-0.81* (0.38)	-0.52* (0.26)	omitted	omitted
g	16.61 (13.47)	28.93*** (8.78)	omitted	omitted
l	-5.15 (23.39)	-0.73 (21.02)	-6.70 (32.60)	-6.70 (32.60)
S.D. E		0.61 (1.23)		omitted
S.D. A		-1.22 (2.32)		omitted
S.D. C		-7.63*** (2.27)		omitted
S.D. N		3.20* (1.61)		omitted
S.D. O		-2.29 (1.66)		omitted
L.L.	-470.91	-460.95	-456.28	-456.28

Notes: *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The robust standard errors are in parentheses. Omitted due to exact collinearity. *L.L.* stands for log-likelihood.

Table 4: Regression results of belief-based error

	Pooled OLS		Fixed Effects	
	Model 3	**Model 4**	**Model 3'**	**Model 4'**
Intercept	8.02	24.01	15.54	15.54
	(10.17)	(14.50)	(9.44)	(9.44)
D^T	-5.70*	-6.92**	-6.39*	-6.39*
	(3.23)	(3.18)	(3.08)	(3.08)
K	-0.13***	-0.14***	-0.15***	-0.15***
	(0.03)	(0.02)	(0.03)	(0.03)
D^T*K	0.07	0.09	0.09*	0.09*
	(0.05)	(0.05)	(0.04)	(0.04)
D^p	0.19	-1.19	-1.41	-1.41
	(2.41)	(2.54)	(2.71)	(2.71)
m	-0.39	-0.23	omitted	omitted
	(0.23)	(0.23)		
g	5.31	11.26	omitted	omitted
	(8.33)	(7.09)		
l	-1.06	0.62	-16.14	-16.14
	(15.06)	(12.71)	(19.09)	(19.09)
S.D. E		0.68		omitted
		(0.74)		
S.D. A		-0.96		omitted
		(1.88)		
S.D. C		-4.99***		omitted
		(1.53)		
S.D. N		2.14*		omitted
		(1.18)		
S.D. O		-1.88		omitted
		(1.47)		
L.L.	-419.99	-428.06	-412.82	-412.82

Notes: *p < 0.1, ** p < 0.05, *** p < 0.01. The robust standard errors are in parentheses. Omitted due to exact collinearity. *L.L.* stands for log-likelihood.

Table 5: The round fixed effects (α_t) and regression results of their trend

	Model 1'	Model 2'	Model 3'	Model 4'
Round 1	23.44		18.82	
Round 2	20.43		19.05	
Round 3	24.88		18.73	
Round 4	17.92		13.44	
Round 5	24.65		15.05	
Round 6	21.79		15.91	
Round 7	19.92		14.30	
Round 8	21.74		18.59	
Intercept	22.76***		18.46***	
	(1.99)		(1.78)	
Slope	-0.20		-0.38	
	(0.39)		(0.35)	
R^2	0.04		0.16	

Notes: *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The robust standard errors are in parentheses.

5.1 Public Information

In Tables 3 and 4, we observed that the coefficients of D^p for the *fundamental-based* error and *belief-based* error have a non-trivial influence on the pricing error albeit in the opposite direction. The average fundamental-based error of positive public information was 17.49% higher than that of the negative case in Model 1' and Model 2'. On the contrary, the positive case caused the average belief-based error to be -1.41% lower than the negative case in Model 3' and Model 4'. This evidence might show that subjects take *public information* to form beliefs and that positive public information has a greater influence than negative public information. In other words, public information first affects the subjects' beliefs and then the market price. By comparing the fundamental-based and belief-based errors, it is found that market price reacts toward the direction of public information but less than the magnitude of the mean belief, which represents the subjects' original thoughts.

5.2 Number of Traders

The effect of the number of traders was significant. Pricing error can be reduced by 0.81% (Model 1) and 0.52% (Model 2) per participant in terms of the fundamental-based error and 0.39% (Model 3) and 0.23% (Model 4) in terms of the belief-based error. Since participants carry independently observed information, the increase in the number of traders can be beneficial to the market's aggregate information. Please note that this result is based on our sampling observation; the actual relationship may, however, not take a linear form between the number of traders and price efficiency (Gresik and Satterthwaite 1989).

5.3   Gender Difference

As can be seen in Table 2, the gender ratio does not remain constant, with the actual male ratio ranging from 31% to 88%. Tables 3 and 4 show that the effect of the gender ratio on price efficiency has disappeared, except for Model 2. Unlike political contracts of the IEM, where male participants demonstrate more aspiration in participation and outward preference expression, the effect of the gender ratio in terms of the individual preference for a specific contract is ignored in our experimental market. This might be because subjects are motivated by monetary reward; their goal is only to seek profitable opportunities as much as possible.

5.4   Limit Order Ratio

Table 2 shows that the mean limit order ratio per round was 0.51, with a standard deviation of 0.07. Since the trading period was restricted, bidding strategies were also conditional on the time constraint. Most participants applied both market and limit orders as much as they could in order to seize profitable opportunities. As such it is not surprising that in Tables 3 and 4 the information regarding the limit order ratio did not improve market efficiency.

5.5   Heterogeneity of Personality Traits

According to the regression results of Model 2 and Model 4 in Tables 3 and 4, the influence of personality traits was concentrated on two aspects, which were conscientiousness, and neuroticism. The coefficients of these variables were significant and consistent in the fundamental- and belief-based models. Among the two traits, the heterogeneity of conscientiousness was the only one that improved pricing efficiency. This result implies that the percentage of *efficient* and *organized* traders in the market is important. The heterogeneity of neuroticism might undermine market efficiency possibly because *sensitive* or *nervous* individuals might find it challenging to make or accept offers based on their information or belief.

5.6   The Degree of Experience with Repeated Trading

From Table 5, we observe that the trend of the fixed effects is declining, being higher in the first few rounds, which suggests that subjects can become familiar with the trading mechanism. As a result, this may improve the efficiency of the experimental prediction market. This phenomenon is generally called *the experience effect*. From Table 3 to Table 5, the results of the pooled OLS are coherent with those of the fixed effects model, implying that our analysis is robust and exonerated from the experience effect. The fixed effects model can be regarded as a robustness check for the pooled OLS after considering the experience effect. Table 5 shows that the slopes are negative but insignificant in both fittings.

6. Conclusion

This study discusses whether the prediction market can achieve efficiency when participants have heterogeneous beliefs and personality traits. Through a series of experiments under the framework of a double auction futures market, it is found that as a medium to communicate and aggregate individual information, the market is hardly efficient even if manipulators are removed. A number of studies have shown that efficiency might be compromised by a number of factors (see Ottaviani and Sørensen, 2007). In this paper, we try to measure the influence of possible factors that cause market failure in the hope of increasing efficiency in the prediction market.

Based on our analysis of the experimental results, there are five possible causes of market inefficiency. First, price is largely influenced by *public information*. Regardless of initial beliefs, participants tend to move in the same direction as what has been disclosed publicly. Second, the larger the number of participants, the more accurate the prediction market can be. Each individual owns a certain amount of information. Through transactions, such information is naturally shared with others, regardless of whether he/she likes it or not. Third, the gender ratio might not be able to improve market efficiency in an experimental market mainly driven by monetary reward. Fourth, no significant bidding behavior can affect market efficiency. Last but not least, personality is found to be a deterministic factor of market efficiency. Participants may hold different views even though they are exposed to identical information. To summarize, two out of the above five personality traits might have some bottom up influence on prices and henceforth the efficiency of the prediction market.

This study demonstrates how individual beliefs can shape or bias the market through aggregation. The relationship between heterogeneous personality and market joint behaviors has also been revealed. This discussion is unprecedented in the experimental prediction market and deserves to receive more attention in future studies.

References

[1] Abdellaoui, M. (2000). Parameter-free elicitation of utilities and probability weighting functions. Management Science, 46, 1497–1512.

[2] Abdellaoui, M., Vossmann, F. and Weber, M. (2005). Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. Management Science, 51, 1384–1399.

[3] Ali, M. M. (1977). Probability and utility estimates for racetrack bettors. Journal of Political Economy, 85, 803–15.

[4] Ballinger, T. P., Hudson, E., Karkoviata, L. and Wilcox, N. T. (2011). Saving behavior and cognitive abilities. Experimental Economics, 14, 349–374.

[5] Berg, J., Nelson, F. and Rietz, T. (2008). Prediction market accuracy in the long run. International Journal of Forecasting, 24, 285–300.

[6] Camerer, C. F. and Ho, T. H. (1994). Nonlinear weighting of probabilities and violations of the betweenness axiom. Journal of Risk and Uncertainty, 8, 167-196.

[7] Christiansen, J. D. (2007). Prediction markets: practical experiments in small markets and behaviours observed. The Journal of Prediction Markets, 1, 17–41.

[8] Costa, P. T. and McCrae Jr., R. R. (1985). The NEO personality inventory manual, Odessa, FL: Psychological Assessment Resources.

[9] Croson, R. and Gneezy, U. (2009). Gender differences in preferences. Journal of Economic Literature, 47, 448–474.

[10] Deck, C., Lin, S. and Porter, D. (2013). Affecting policy by manipulating prediction markets: experimental evidence. Journal of Economic Behavior & Organization, 85, 48–62.

[11] Dreber, A., von Essen, E. and Ranehill, E. (2011). Outrunning the gender gap—boys and girls compete equally. Experimental Economics, 14, 567–582.

[12] Ehmke, M. D. T. (2006). Dissertation abstract: The influence of culture on economic behavior with applications to food and the environment. Experimental Economics, 9, 167–168.

[13] Figlewski, S. (1978). Market 'efficiency' in a market with heterogeneous information. Journal of Political Economy, 86, 581–597.

[14] Figlewski, S. (1979). Subjective information and market efficiency in a betting market. The Journal of Political Economy, 87, 75–88.

[15] Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics, 10, 171–178.

[16] Forsythe, R., Nelson, F., Neumann, G. and Wright, J. (1992). Anatomy of an experimental political stock market. The American Economic Review, 82, 1142–1161.

[17] Forsythe, R., Rietz, T. and Ross, T. (1999). Wishes, expectations and actions: a survey on price formation in election stock markets. Journal of Economic Behavior & Organization, 39, 83–110

[18] Gjerstad, S. (2005). Risk aversion, beliefs, and prediction market equilibrium. Unpublished Manuscript, Economic Science Laboratory, University of Arizona.

[19] Gresik, T. A. and Satterthwaite, M. A. (1989). The rate at which a simple market converges to efficiency as the number of traders increases: an asymptotic result for optimal trading mechanisms. Journal of Economic Theory, 48, 304–332.

[20] Griffith, R. M. (1949). Odds adjustments by American horse-race bettors. The American Journal of Psychology, 62, 290–294.

[21] Hahn, R. W. and Tetlock, P. C. (2006). Information markets: a new way of making decisions, AEI-Brookings Joint Center for Regulatory Studies.

[22] Hanson, R. D. (2006). Designing real terrorism futures. Public Choice, 128, 257–274.

[23] Haruvy, E., Lahav, Y. and Noussair, C. N. (2007). Traders' expectations in asset markets: experimental evidence. American Economic Review, 97, 1901–1920.

[24] Hayek, F. (1945). The use of knowledge in society. American Economic Review, 35, 519–530.

[25] Hazen, T. L. (1987). Volatility and market inefficiency: a commentary on the effects of options, futures, and risk arbitrage on the stock market. Washington and Lee Law Review, 44, 789–805.

[26] Isaacs, R. (1953). Optimal horse race bets. American Mathematical Monthly, 60, 310–315.

[27] Knetsch, J., Tang, F. F. and Thaler, R. (2001). The endowment effect and repeated market trials: is the Vickrey auction demand revealing? Experimental Economics, 4, 257–269.

[28] Loomes, G., Starmer, C. and Sugden, R. (2003). Do anomalies disappear in repeated markets? Economic Journal, 113, C153–C166.

[29] Maloney, M. T. and Mulherin, J. H. (2003). The complexity of price discovery in an efficient market: the stock market reaction to the Challenger crash. Journal of Corporate Finance, 9, 453–479.

[30] Oechssler, J. (2012). Finitely repeated games with social preferences. Experimental Economics, doi:10.1007/s10683-012-9336-6.

[31] Oliven, K. and Rietz, T. A. (2004). Suckers are born but markets are made: individual rationality, arbitrage, and market efficiency on an electronic futures market. Management Science, 50, 336–351.

[32] Oosterbeek, H., Sloof, R. and Van De Kuilen, G. (2004). Cultural differences in ultimatum game experiments: evidence from a meta-analysis. Experimental Economics, 7, 171–188.

[33] Ottaviani, M. and Sørensen, P. N. (2007). Outcome manipulation in corporate prediction markets. Journal of the European Economic Association, 5, 554–563.

[34] Ottaviani, M. and Sørensen, P. N. (2010). Noise, information, and the favorite-longshot bias in parimutuel predictions. American Economic Journal: Microeconomics, 2, 58–85.

[35] Roll, R. (1984). Orange juice and weather. American Economic Review, 74, 861–880.

[36] Rosenberg, B., Reid, K. and Lanstein, R. (1985). Persuasive evidence of market inefficiency. Journal of Portfolio Management, 11, 9–16.

[37] Snowberg, E., Wolfers, J. and Zitzewitz, E. (2007). Partisan impacts on the economy: evidence from prediction markets and close elections. Quarterly Journal of Economics, 122, 807–829.

[38] Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. Journal of Risk and Uncertainty, 32, 101–130.

[39] Tufano, F. (2010). Are 'true' preferences revealed in repeated markets? an experimental demonstration of context-dependent valuations. Experimental Economics, 13, 1–13.

[40] Tversky, A. and Fox, C. (1995). Weighing risk and uncertainty. Psychological Review, 102, 269–283.

[41] Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. Journal of Risk and Uncertainty, 5, 297–323.

[42] Visser, M. S. and Roelofs, M. R. (2011). Heterogeneous preferences for altruism: gender and personality, social status, giving and taking. Experimental Economics, 14, 490–506.

[43] Weitzman, M. (1965). Utility analysis and group behavior: an empirical study. Journal of Political Economy, 73, 18–26.

[44] Wolfers, J. and Zitzewitz, E. (2004). Prediction markets. The Journal of Economic Perspectives, 18, 107–126.

[45] Wu, G. and Gonzalez, R. (1996). Curvature of the probability weighting function. Management Science, 42, 1676–1690.