# Spatial Modeling of Agent-Based Prediction Markets: Role of Individuals

Bin-Tzong Chie[1] and Shu-Heng Chen[2(✉)]

[1] Department of Industrial Economics, Tamkang University,
Tamshui, New Taipei City 251, Taiwan
chie@mail.tku.edu.tw

[2] Department of Economics, AIECON Research Center,
National Chengchi University, Taipei 116, Taiwan
chen.shuheng@gmail.com

**Abstract.** In this paper, we extend the spatial agent-based prediction market proposed by Yu and Chen at MABS 2011 into a spatial model in which agents choose their community (neighbors) by following Schelling's proximity model. This extended model generalizes the spatial configuration of the original model and enables us to examine the validity of the Hayek hypothesis when the information distribution is determined by clusters of agents with heterogeneous identities. Specifically, we examine the role of the toleration capacity, the key parameter in the Schelling model, which generates the clusters of agents with different sizes, and the role of exploration capacity which determines how well an agent is informed about his local surroundings. We find that after taking into account market activity and price volatility, both the toleration capacity and exploration capacity have a positive effect on the prediction accuracy and enhance information polling and the information aggregation of markets. The results obtained in this agent-based simulation, therefore, add a qualification to the well-known Hayek hypothesis and point to the significance of individuals in information aggregation.

## 1 Motivation and Introduction

How accurately the prediction market can predict, up to the present, is basically an empirical issue. However, empirical studies per se cannot articulate why sometimes the market for some events performed extremely well and sometimes it did not [2]. While there are a number of studies trying to identify the factors contributing to its successes or failures, the explanations supporting the found causal links remain very verbal and informal, and a rigorous mechanism has not been explicitly spelled out. This is partially due to the limited analytical tractability of the prediction markets which operate in practice. In this article, we argue that, the spatial configuration, i.e., the distribution of information over agents, situated in different places, can matter for the prediction accuracy of the prediction markets. However, since the usual analytical model cannot effectively deal with these geographical variables, an agent-based spatial model of prediction markets is proposed to address the geographical significance. To begin with

F. Grimaldo and E. Norling (Eds.): MABS 2014, LNAI 9002, pp. 197–212, 2015.
DOI: 10.1007/978-3-319-14627-0_14

this line of research, our model is tailored to the future events related to political elections only, normally known as the *political futures*. In other words, we shall show how geographical factors can be part of the functioning of the prediction accuracy of the political futures markets.

The rest of the paper is organized as follows. Section 2 introduces our proposed spatial agent-based prediction markets and the two essential ingredients in the model, namely, toleration capacity and exploration capacity. Section 3 discusses the design of our simulation and shows the simulation results. Section 4 gives the concluding remarks.

## 2 The Model

### 2.1 The Market

**Network-Based Formation of Expectations and Reservation Prices.** Our first step is to make the social network explicit (Sect. 2.2). Through the given social network, agents disseminate and acquire the information and form their expectations of the future election outcomes, upon which their decisions on bids and asks are based. We assume that, to form an expectation regarding the election outcome, all agents use the sample average as the estimate, and the sample available for each agent is identical to the set of all his connecting agents (to be defined later). In other words, by using the sample proportion of the connecting agents supporting each political candidate, the agent forms his expectations about the share of the vote of each candidate. This estimated share becomes the *reservation price* held by the agents. To make this point precise, let $\hat{p}_{i,j}$ be the subjective estimation of agent $i$ regarding the share of the votes attributed to candidate $j$, and $b_{i,j}$ be the reservation price that agent $i$ holds for the futures related to the vote share of candidate $j$. Then

$$b_{i,j} = \hat{p}_{i,j} = \frac{\#\{k : k \in N_i \cap V_j\}}{\#N_i}, \quad i = 1, 2, ..., N, \quad j = 1, ..., m, \tag{1}$$

where $N_i$ is the set of agent $i$'s connecting agents (to be defined later), and $V_j$ is the set of voters who support candidate $j$. By (1), if the estimated share of the votes of Candidate A is 60 %, then the reservation price of the future contract for the share of votes of Candidate A is 60 cents. With this reservation price, the agent would not accept any bids which are lower than 60 or any asks which are higher than 60.

**Bidding and Asking Strategy.** In fact, following most agent-based prediction markets [5,9], we assume that all agents are *zero-intelligent agents* (the entropy-maximizing agent) in the sense that the agent will bid or ask randomly with the constraint of making no expected loss [1,4]. Therefore, his bid $p_{b,i,j}$ will be uniformly sampled from the interval between the floor, which is zero cents, and the reservation price $b_{i,j}$, and his ask $p_{a,i,j}$ will be uniformly sampled from the

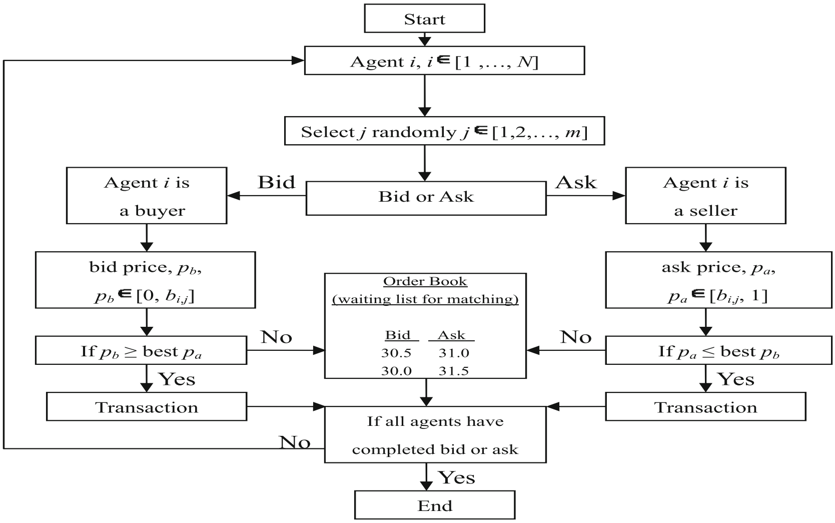

Start
Agent $i$, $i \in [1, \ldots, N]$
Select $j$ randomly $j \in [1,2,\ldots, m]$
Bid or Ask
Bid
Ask
Agent $i$ is a buyer
Agent $i$ is a seller
bid price, $p_b$, $p_b \in [0, b_{i,j}]$
ask price, $p_a$, $p_a \in [b_{i,j}, 1]$
Order Book (waiting list for matching)
Bid Ask
30.5 31.0
30.0 31.5
If $p_b \geq$ best $p_a$
If $p_a \leq$ best $p_b$
No
No
Yes
Yes
Transaction
Transaction
If all agents have completed bid or ask
No
Yes
End

**Fig. 1.** The flowchart of the order-book driven prediction market

interval between his reservation price and the ceiling, which is one dollar, as shown in Eq. (2).

$$p_{b,i,j} \sim U[0, b_{i,j}], \quad p_{a,i,j} \sim U[b_{i,j}, 1], \quad i = 1, 2, ..., N, \quad j = 1, 2, ..., m. \qquad (2)$$

**Trading Mechanism.** The trading mechanism adopted to run the market is continuous double-auction, the one frequently used in experimental economics to test the Hayek hypothesis [7]. As shown in Fig. 1, our agent-based prediction market starts from a random draw of the agents. Each agent shall be drawn exactly once; in other words, the draw proceeds in a sampling-without-replacement manner. When agent $i$ is drawn, he will be randomly placed into one of the $m$ markets and will be equally likely to be assigned either a buyer position or a seller position. He will then submit a bid if he is a buyer and submit an ask if he is a seller. His bid or ask will be placed in the order book. A match happens if either his bid ($p_{b,i,j}$) is greater than the remaining lowest ask (best$p_a$) in the order book or his ask ($p_{a,i,j}$) is lower than the remaining highest bid (best$p_b$). The transaction price will then be determined as best$p_a$ if the former applies or as best$p_b$ if the latter applies.

## 2.2 Geographical Distribution of Agents

The social networks considered in this paper are generated from the *Schelling segregation model* [6], in which the location of agents is determined by their *toleration capacity* for agents with different political identities. In other words, we replace the ethnic heterogeneity of agents in the original Schelling model with

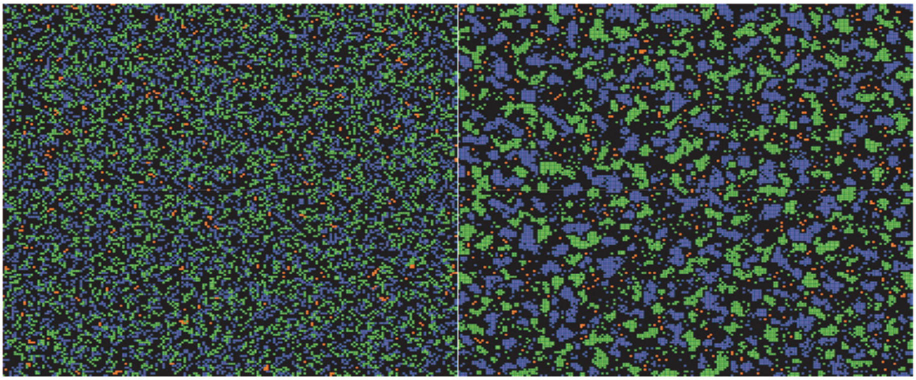


**Fig. 2.** Geographical distribution of voters and their political identity. Both panels are the converged configurations using $v_1 = 45.63\,\%$ (green), $v_2 = 51.60\,\%$ (blue), $v_3 = 2.77\,\%$ (orange), $N = 13,454$, and $G$ (number of grids) $= 193 \times 193$. The black grids denote the unoccupied cells, and the colored grids denote the occupied cells. The number of occupied cells and the number of unoccupied cells are determined in such a way that the resultant population density is close to 36 % (see Table 1). The two panels differ in terms of the toleration capacity: on the left, $s = 0.75$, and, on the right, $s = 0.25$. (Color figure online)

their *political identity* ($j = 1, 2, ..., m$). Agents tend to reside in the place which is surrounded by neighbors with the same political identity. Their toleration of neighbors with different political identities is characterized by the parameter, *toleration capacity* ($s$). If the ratio of neighbors with different political identities is larger than this threshold $s$, they tend to move to a close place which their toleration capacity can handle. This migration process will be iterated until it converges to a fixed configuration. We then use the resultant configuration to represent the geographical distribution of residents with different political identities.

Apart from the toleration capacity, an additional parameter of Schelling's segregation model is the demographical structure characterized by the percentage of agents of various political identities. Denote them by $v_j$ ($j = 1, 2, ..., m$).

$$v_j = \frac{\#(V_j)}{N}, j = 1, 2, ..., m, \tag{3}$$

where $N$ is the total number of agents.

Figure 2 demonstrates a geographical distribution of political identities. In this specific example, there are a total of 13,454 agents, distributed on a checkerboard with $193 \times 193$ grids, i.e., with a population density of 36.12 %, and $m = 3$ (three candidates or three political parties): $v_1 = 45.63\,\%$, $v_2 = 51.60\,\%$, and $v_3 = 2.77\,\%$. Agents with the three political identities are denoted by the green

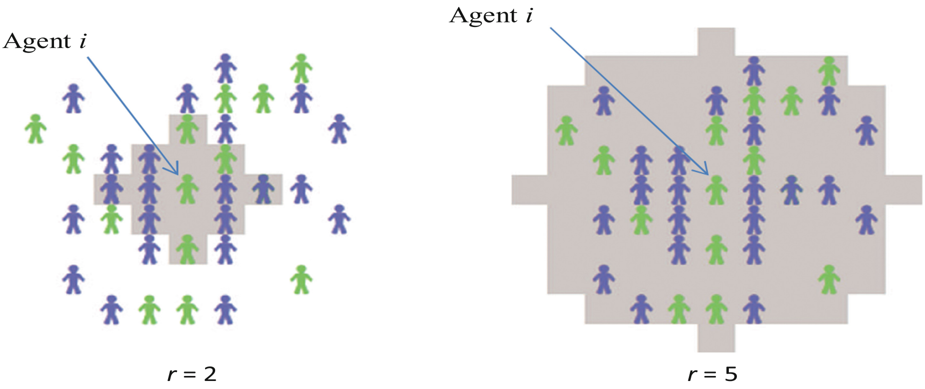


**Fig. 3.** The von Neumann Neighborhood with a radius of 2 (left) and 5 (right). The above figures show the von Neumann neighborhood of agent $i$, as pointed to by an arrow. The left panel is a neighborhood with a radius of 2, whereas the right panel is a neighborhood with a radius of 5. (Color figure online)

($j = 1$), blue ($j = 2$), and the orange color ($j = 3$), respectively.[1] What is demonstrated in Fig. 2 are, therefore, two of the converged configurations of agents who followed the Schelling rule of migration. The one on the left is the one corresponding to a toleration capacity of 0.75, and the one on the right is the one corresponding to a toleration capacity of 0.25.

### 2.3 Exploration Capacity

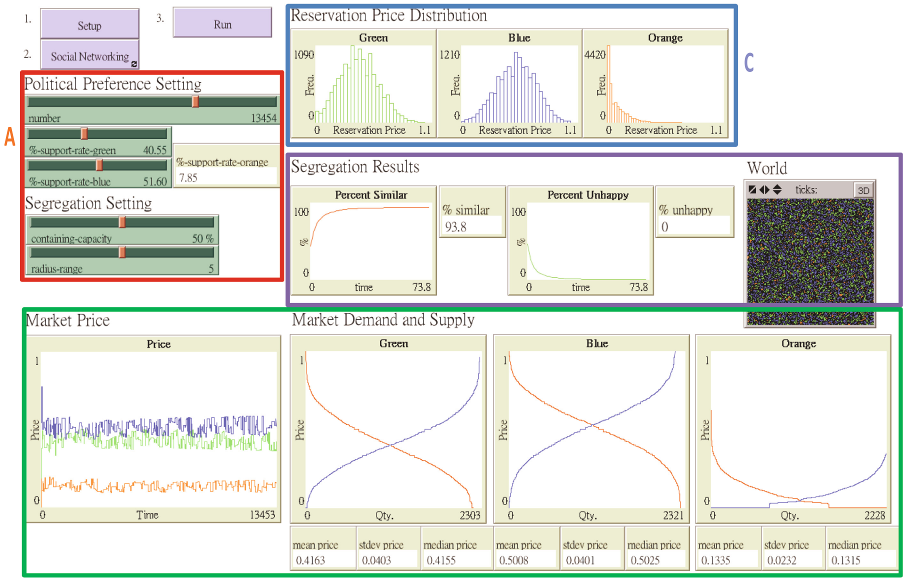
For each agent, his information supplier, i.e., his set of connecting agents, is determined by a von Neumann neighborhood with a given radius ($r$). This is shown in Fig. 3. As shown in Eq. (1), agents are assumed to know the political identities of all of their connecting agents in the neighborhood (agents in the gray area), and they use this sample (local information) to estimate the share of the votes for each candidate. The radius, $r$, can be interpreted as the information exploration capacity of the agent. The larger the radius the larger the sample, and hence the less biased and the better the estimation. In this article, we assume that agents are homogeneous with respect to this capacity but would like to examine how this parameter may affect the emergent market performance.

### 2.4 Programming with NetLogo

The above-mentioned spatial agent-based prediction market is programmed with NetLogo 5.0.3 and is available from the OpenABM website[2]. Figure 4 shows a familiar NetLogo display of running this program.

[1] These parameter values are based on the 2012 Presidential Election in Taiwan. Based on the 2012 Presidential election outcome, the DPP candidate (colored in green) won a share of 45.63 % of the vote, the KMT candidate (colored in blue) won a share of 51.60 %, and the PFP candidate (colored in orange) won a share of 2.77 %.

[2] http://www.openabm.org/model/3764/.



number 13454
%-support-rate-green 40.55
%-support-rate-blue 51.60
%-support-rate-orange 7.85
containing-capacity 50 %
radius-range 5

% similar 93.8
% unhappy 0

| mean price | stdev price | median price | mean price | stdev price | median price | mean price | stdev price | median price |
|---|---|---|---|---|---|---|---|---|
| 0.4163 | 0.0403 | 0.4155 | 0.5008 | 0.0401 | 0.5025 | 0.1335 | 0.0232 | 0.1315 |

**Fig. 4.** Display of the NetLogo program (Color figure online).

In Fig. 4, the upper left panel (panel A) gives the user-supplied control parameters: $N = 13,454$, $v_1 = 40.55\,\%$ (green), $v_2 = 51.60\,\%$ (blue), $v_3 = 7.85\,\%$ (orange), $s = 0.50$ (50 %) and $r = 5$. The diagram shown in the right middle panel (panel B) is the converged configuration using the Schelling rule with $s = 0.5$. With a radius of 5, we can have the price expectations (reservation prices) of all three futures for all agents, i.e., $b_{i,j}$ ($i = 1, ..., 13454$, and $j = 1, 2, 3$). What is shown in the right upper panel (panel C) of the figure are the three histograms of the reservation prices corresponding to the green, blue and orange party, respectively. The basic statistics, including the mean, the median and the standard deviation, are shown in the very bottom of the figure (panel D). There we can see that the mean and median for the green candidate are 0.4163 and 0.4155, which is a one-point upward bias away from the true value of 0.4055. In addition, for the blue candidate, these two statistics are 0.5008 and 0.5025, which is a one-point downward bias away from the true value of 0.5160. Maybe the worst case is the market for the orange candidate. The two corresponding statistics are 0.1335 and 0.1315, almost two times larger than the true value of 0.0785. Our research question is then, to what extent, this specific network topology may affect the accuracy of the prediction market or the political futures market in our case.

From the histogram, we can further derive the aggregate willingness to buy (when the price is below the reservation price)

$$Q_j^D(p) = \#\{i : b_{i,j} > p\}, \tag{4}$$

and the aggregate willingness to sell (when the price is above the reservation price)

$$Q_j^S(p) = \#\{i : b_{i,j} < p\} \tag{5}$$

i.e., the demand curve ($Q_j^D$) and the supply curve ($Q_j^S$).

The demand and supply curves of the three markets are shown in the lower middle and right panels (panel D). Then through the random draws of the agents and their reservation price, the order book for each market is formed, and the corresponding transaction price is generated as the time series shown in the lower left panel of the figure.

# 3 Simulation

## 3.1 Simulation Design

The main focus of this paper is to understand how the information aggregation can be affected by how it is distributed through the two control parameters, namely, toleration capacity ($s$) and exploration capacity ($r$). In fact, we believe that these two parameters, to some extent, characterize the quality of voters, their cultural backgrounds, sociability, and openness. None of these attributes has been mentioned in the original article of the Hayek hypothesis [3]. Presumably, they are all irrelevant or insignificant. This paper is purported to revisit this hypothesis from a cultural and social-psychological aspect.

Given this focus, most parameters should be held constant throughout the simulation, and include $N$, $m$, $d$, and $G$ (Table 1). Nonetheless, to make the choice of these parameters not entirely arbitrary and to clothe them with some empirical flavor, we use the real data from Taiwan to suggest some reasonable values of these parameters. According to the 2010 demographic census data in Taiwan, the number of qualified voters in the 2012 presidential election was 13,453,305. By scaling down the number of people by 1,000 times, there are 13,454 agents. Hence, $N$ is set to 13,454. In addition, by considering the population density of Taiwan, $d$ is set to 36.12 %, which implies that we need to have a grid size of 193 × 193.[3] Hence, $G$ is also determined. As to the number of candidates, in the most recent Presidential election in Taiwan, held in the year 2012, there were three major political parties and hence three major candidates. Hence, $m$ is set to 3. This finishes the description of constant parameters in Table 1.

The rest of the prediction market is characterized by four major parameters, $s$, $r$, $v_1$, and $v_3$. We first give a range for each of these parameters; each design can be regarded as a three-tuple randomly selected from this range. For $s$, we consider a range from a low toleration capacity (0.26) to a high toleration capacity (0.75), with an increment of 0.01. The exploration capacity ($r$), it starts with a minimum of 2, and ends with a maximum of 6. Finally, for $v_i$, considering the practice of Taiwan politics, we fix the share of the votes for the small party, i.e., 3 %, and

[3] Taiwan's population density is around 630 people per square kilometer. If we only consider the number of qualified voters, and not the entire population size, then the population density is approximately 372 per square kilometer. By assuming that one square kilometer is roughly equal to 32 × 32 grids, we can then figure out the required $d$ (36.12 %) and the number of grids (193 × 193).

**Table 1.** Tableau of control parameters

| Parameter | Description | Value |
|---|---|---|
| $m$ | The number of candidates | 3 |
| $s$ | Containing capacity | 0.25, 0.26, ..., 0.75 |
| $r$ | Exploration capacity (Radius) | 2, 3, ..., 6 |
| $v_1$ | Vote share of Green candidate | 18, 19, ..., 47 |
| $v_2$ | Vote share of Blue candidate | 100 - $v_1$ - 3 |
| $v_3$ | Vote share of Orange candidate | 3 |
| $N$ | Number of agents | 13,454 |
| $d$ | Population density | 36.12 % |
| $G$ | Grid size | 193 × 193 |
| $R$ | Simulation runs | 50 |

then allow the other two major parties to vary in opposite directions. Again, from an empirical consideration, the range of $v_1$ is set from 18 to 47, and then $v_2$ takes the rest. We then randomly generate 1,000 designs, and each design is run 50 times. To sum up, we have

$$Design_k \equiv \{s_k, r_k, v_{1,k}\}, k = 1, 2, ..., 1000, \tag{6}$$

where

$$s_k \sim U[0.26, 0.75], r_k \sim U[2, 3, 4, 5, 6], v_{1,k} \sim U[19, 47]. \tag{7}$$

The random design described above allows us to have enough observations to examine the effect of these two parameters on the emergent market performance.

### 3.2 Basic Results

Table 2 shows that the results for each design look like. Notice that we do not present all of them; otherwise, the table would be 1,000 rows long, since we have a total of 1,000 designs. Each row starts with parameters characterizing the design, namely, $s, r, v_1, v_2$, and $v_3$, followed by the key summary statistics of each design, including the mean price, trading volume, and volatility (standard deviation of the price) of each future. Since each design has been run 50 times, all these statistics are the averages taken over 50 runs. For the mean price, we first take the average of the price series for each run (Eq. 9), and take the average of the average over these 50 runs (Eq. 8).

$$\bar{p}_j = \frac{\sum_{l=1}^{50} \bar{p}_{j,l}}{50}, \quad j = 1, 2, 3, \quad l = 1, 2, ..., 50, \tag{8}$$

where

$$\bar{p}_{j,l} = \frac{\sum_{t_{j,l}=1}^{T_{j,l}} p_{j,l}(t_{j,l})}{T_{j,l}}, j = 1, 2, 3, \quad l = 1, 2, ..., 50, \tag{9}$$

**Table 2.** Simulation input and output table

| $s$ | $r$ | $v_1$ | $v_2$ | $v_3$ | $\bar{p}_1$ | $\bar{p}_2$ | $\bar{p}_3$ | $Vol_1$ | $Vol_2$ | $Vol_3$ | $\sigma_1$ | $\sigma_2$ | $\sigma_3$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.26 | 2 | 34 | 63 | 3 | 30.66 | 66.38 | 2.95 | 736.0 | 762.0 | 111.9 | 0.1208 | 0.1289 | 0.0217 |
| 0.26 | 2 | 39 | 58 | 3 | 36.42 | 60.71 | 2.87 | 794.8 | 814.3 | 109.2 | 0.1336 | 0.1389 | 0.0204 |
| 0.26 | 2 | 41 | 56 | 3 | 38.97 | 58.19 | 2.84 | 806.5 | 832.9 | 109.9 | 0.1367 | 0.1426 | 0.0201 |
| 0.26 | 2 | 46 | 51 | 3 | 45.54 | 51.61 | 2.85 | 842.6 | 845.6 | 111.9 | 0.1423 | 0.1449 | 0.0196 |
| . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 0.75 | 6 | 42 | 55 | 3 | 41.25 | 51.88 | 6.87 | 236.2 | 238.4 | 147.3 | 0.0369 | 0.0428 | 0.0131 |
| 0.75 | 6 | 44 | 53 | 3 | 42.95 | 50.19 | 6.86 | 235.5 | 234.4 | 152.0 | 0.0378 | 0.0417 | 0.0134 |
| 0.75 | 6 | 45 | 52 | 3 | 43.70 | 49.43 | 6.87 | 232.3 | 238.4 | 150.2 | 0.0386 | 0.0416 | 0.0132 |

and $T_{j,l}$ are the transaction times of future $j$ in the $l$th run.

These three figures, $\bar{p}_j$ $(j = 1, 2, 3)$ are shown in the first three columns of the right panel of Table 2.[4] The next three columns, $Vol_j$ $(j = 1, 2, 3)$ are the average of the trading volume over the 50 runs, and likewise for the price volatility.

$$\sigma_j = \frac{\sum_{i=1}^{50} \sigma_{j,l}}{50}, \quad j = 1, 2, 3; \quad l = 1, 2, ..., 50, \tag{12}$$

where $\sigma_{j,l}$ is the standard deviation of the price of the $j$th future in the $l$th run. Table 2, therefore, provides us the basic input (the left panel) and output (the right panel) correspondence which allows us to address further the effect of the two key parameters, $s$ and $r$, on the prediction accuracy.

Based on Table 2, we shall start with a simple linear regression.

$$Y = f(s, r) + \epsilon = \beta_0 + \beta_1 s + \beta_2 r + \epsilon. \tag{13}$$

The dependent variable $Y$ is the prediction accuracy based on the chosen error functions. In this paper, we shall use $\bar{p}_j$ as the key predictor of $v_j$ and consider the following four error measures frequently used in the literature.

[4] We assume that the non-arbitrage condition is always satisfied, i.e.,

$$\sum_{j=1}^{3} \bar{p}_{j,l} \times 100 = 100, \forall l \tag{10}$$

However, if the above equality is violated, then we shall rescale our mean price as follows,

$$\bar{p}_{j,l}^{adj} = \frac{\bar{p}_{j,l}}{\sum_{j=1}^{3} \bar{p}_{j,l}} \times 100, \tag{11}$$

and use the re-scaled price $\bar{p}_{j,l}^{adj}$ to replace $\bar{p}_{j,l}$ in Eq. (8).

1. Mean Absolute Percentage Error (MAPE)

$$Y_1 = MAPE = \frac{\sum_{j=1}^{m} \mid \bar{p}_j - v_j \mid /v_j}{m} \tag{14}$$

2. Root Mean Square Error (RMSE):

$$Y_2 = RMSE = \sqrt{\frac{\sum_{j=1}^{m} (\bar{p}_j - v_i)^2}{m}} \tag{15}$$

3. Mean Square Error (MSE)

$$Y_3 = MSE = \frac{\sum_{j=1}^{m} (\bar{p}_j - v_j)^2}{m} \tag{16}$$

4. Euclidian Distance (ED)

$$Y_4 = ED = \sqrt{\sum_{j=1}^{m} (\bar{p}_j - v_j)^2} \tag{17}$$

The results of the prediction errors over these four error measures are provided in Table 3. Again, this is a simplified modification by only showing the first few and the last few rows. A complete table has 1,000 rows. This table then serves as the basis for running the linear regression (13).

The first regression result is shown in Table 4 (the upper panel). There we find that both $s$ and $r$ have a negative effect on the prediction accuracy, i.e., $\beta_1 > 0$ and $\beta_2 > 0$, and the result is consistent regardless of the measure being employed. This result is somewhat counter intuitive, since one might initially have thought that increasing either the toleration capacity ($s$) or the exploration capacity ($r$) can make individual agents more informative, which in turn may help the information aggregation in the later stage. Nevertheless, this is not the case which we have here, but why? One possible explanation is that when both $s$ and $r$ become larger, depending on the $v_j$, agents are not just better informed, but also more homogeneous in their expectations and reservation prices, which may cause transactions more difficult to happen and make the market less liquid. One such famous example is Tirole's zero-trading theorem [8], i.e., in an extreme case where agents are all perfectly informed, there will be no trade in the market; in other words, the market can predict nothing at all in this situation.

### 3.3 Homogeneity Effect

To see this homogeneity effect, Fig. 5 shows the average trading volume under different vote shares with respect to these two capacities. Three features immediately stand out.

First, there are hump-shaped curves in each sub-diagram with respect to a given exploration capacity (the left panel) or with respect to a given toleration

**Table 3.** Prediction accuracy

| $s$ | $r$ | $v_1$ | $v_2$ | $v_3$ | MAPE | RMSE | MSE | ED |
|---|---|---|---|---|---|---|---|---|
| 0.26 | 2 | 34 | 63 | 3 | 0.0558 | 2.7434 | 7.5260 | 4.7516 |
| 0.26 | 2 | 39 | 58 | 3 | 0.0522 | 2.1605 | 4.6678 | 3.7421 |
| 0.26 | 2 | 41 | 56 | 3 | 0.0472 | 1.7278 | 2.9853 | 2.9926 |
| 0.26 | 2 | 46 | 51 | 3 | 0.0238 | 0.4468 | 0.1996 | 0.7739 |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| 0.75 | 6 | 42 | 55 | 3 | 0.4544 | 2.9006 | 8.4134 | 5.0239 |
| 0.75 | 6 | 44 | 53 | 3 | 0.4551 | 2.8261 | 7.9870 | 4.8950 |
| 0.75 | 6 | 45 | 52 | 3 | 0.4557 | 2.7824 | 7.7415 | 4.8192 |

capacity (the right panel) indicating that the trading volume increases when competition between the major political parties is keen, i.e., the share of the vote of the two major candidates is close.

Second, however, the hump-shaped curve has a tendency to shift down with the *increase* in each of the two capacities. Since the higher the capacities, the more homogeneous is the information received by the agent, the pattern of the shifting-down hump-shared curves indicates that the trading volume goes down with the degree of homogeneity.

Third, the curvature of the hump-shaped curve also decreases with the increase in the toleration capacity (the left panel) or the increase in the exploration capacity (the right panel). For example, when these capacities are higher, such as up to 70 % (for $s$) or up to 6 (for $r$), the hump is flattened out. This indicates that the effect of the uncertainty, measured by the closeness of the two major candidates in their share of the vote, no longer affects the trading volume when voters are homogeneously well-informed. This is not surprising: when voters are homogeneously well-informed, market uncertainty perceived by voters is reduced and hence even a neck-to-neck competition has little effect on the trading volume. To sum up, our analysis above shows that, in addition to the vote share or market uncertainty, the two capacities also affect the trading volume, and they affect it in a downward direction.

The same analysis is further carried out for the price volatility. Figure 6 shows the effect of the two capacities on the average price volatility (Eq. 12). Qualitatively speaking, the result is the same. All three features with regard to the effect of the two capacities remain for the case of the price volatilities. The trading volume (the thickness of the market) with the price volatility is the indicator of a functioning market where information is aggregated and revealed. However, when the degree of homogeneity of traders is high, these functions are adversely affected.

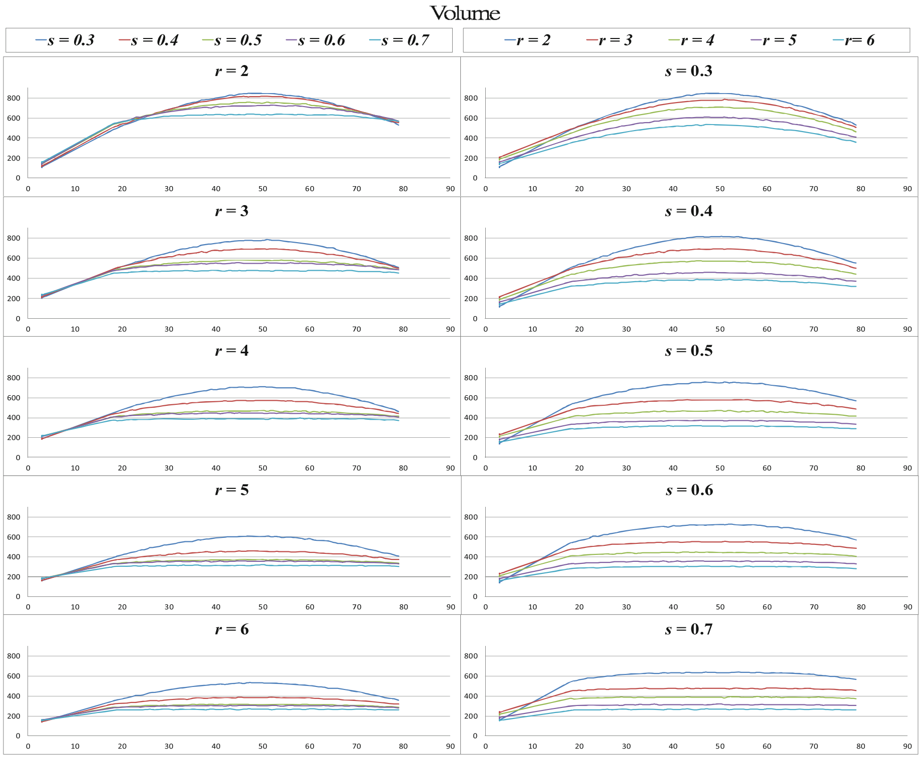


**Fig. 5.** Trading volume, exploration capacity, and toleration capacity. The five sub-diagrams in the left panel are drawn in the way by fixing the exploration capacity ($r$) and examining the effect of the toleration capacity ($s$) on the trading volume. To see the difference, different values of $s$ are colored differently. The five sub-diagams in the right panel are drawn in the way by fixing the toleration capacity ($s$) and then examining the effect of the exploration capacity on the trading volume. Again, to see the difference, different values of $r$ are colored differently.

### 3.4 Conditional Regression

Given the homogeneity effect, it would be desirable to control some market characteristics while running the regression against $s$ and $r$. Therefore, we propose a second linear regression which takes into account the market characteristics. Two usual market characteristics considered in the literature are the trading volume ($Vol$) and the price volatility ($\sigma$). Following this convention, we propose the second linear regression (18).

$$Y = \beta_0 + \beta_1 s + \beta_2 r + \sum_{i=3}^{5} \beta_i Vol_{i-2} + \sum_{i=6}^{8} \beta_i \sigma_{i-5} + \epsilon, \qquad (18)$$

where $Vol_i$ $(i = 1, 2, 3)$ is the trading volume of the $i$th futures, and $\sigma_i$ $(i = 1, 2, 3)$ is the price volatility of the corresponding futures.

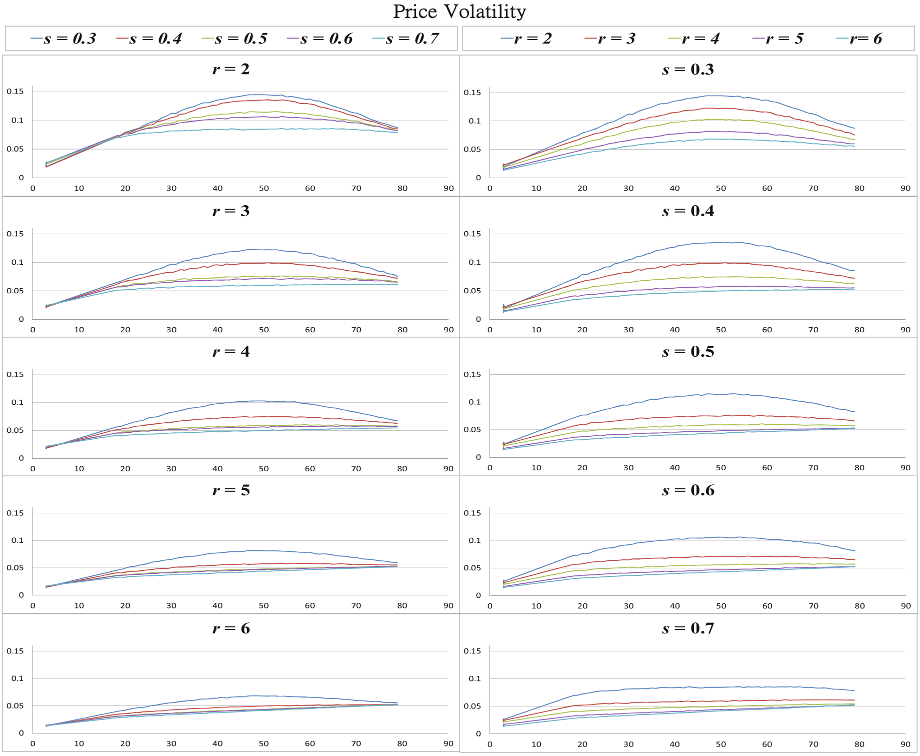


**Fig. 6.** Price volatility, exploration capacity, and toleration capacity. The five sub-diagrams in the left panel are drawn in the way by fixing the exploration capacity ($r$) and examining the effect of the toleration capacity ($s$) on the price volatility. To see the difference, different values of $s$ are colored differently. The five sub-diagrams in the right panel are drawn in the way by fixing the toleration capacity ($s$) and then examining the effect of the exploration capacity on the price volatility. Again, to see the difference, different values of $r$ are colored differently.

Since, as we have seen in Sect. 3.3, the trading volume and the price volatility have already been "polluted" by the two capacities (Figs. 5 and 6), in econometrics, this is what is familiarly known as an *endogeneity problem*. To take care of the endogeneity problem, what we do here is then, first, to run the two auxiliary regressions, one on the trading volume and one on the price volatility, against the two capacities, then, second, to take the residuals as the "cleaned" (filtered) trading volume and volatility. We then use them as independent variables in the market performance regression (18).

The regression results of regression (18) are shown in the lower panel of Table 4. The results show that the inclusion of the market characteristics can improve the coefficient of determination ($R^2$). This result is not difficult to understand. Given the geographical complexity and variability of the two-dimensional lattice, controlling both $s$ and $r$ does not automatically imply the control of the

**Table 4.** Regression results with market characteristics

Regression results without market characteristics

| | $\beta_0$ | $\beta_1$ | $\beta_2$ | | | | | | | $\bar{R}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| MAPE | −0.034 | 0.0032 | 0.073 | | | | | | | 0.4 |
| | 0.0776 | 0.0000 | 0.0000 | | | | | | | |
| RMSE | 0.2987 | 0.0312 | 0.4177 | | | | | | | 0.27 |
| | 0.0826 | 0.0000 | 0.0000 | | | | | | | |
| MSE | −7.086 | 0.2525 | 2.2298 | | | | | | | 0.21 |
| | 0.0000 | 0.0000 | 0.0000 | | | | | | | |
| ED | 0.5174 | 0.054 | 0.7235 | | | | | | | 0.27 |
| | 0.0826 | 0.0000 | 0.0000 | | | | | | | |

Regression results with market characteristics

| | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\beta_6$ | $\beta_7$ | $\beta_8$ | $\bar{R}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| MAPE | 2.668 | −0.013 | −0.171 | 0.001 | −0.002 | 0.003 | −3.066 | 4.566 | 7.222 | 0.98 |
| | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | |
| RMSE | 17.724 | −0.078 | −1.304 | 0.035 | −0.040 | 0.012 | −171.418 | 179.218 | 88.613 | 0.81 |
| | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | |
| MSE | 114.085 | −0.526 | −10.081 | 0.309 | −0.351 | 0.061 | −1473.720 | 1649.300 | 790.791 | 0.78 |
| | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | |
| ED | 30.700 | −0.136 | −2.259 | 0.061 | −0.069 | 0.021 | −296.905 | 310.415 | 153.482 | 0.81 |
| | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | |

The two panels above show the regression results of regression Eqs. (13) and (18), respectively. The former does not include the market characteristics as the independent variables, whereas the latter does. The two regressions were run using different dependent variables (performance criteria). The first column gives the dependent variable used in the respective regression. The results for each regression and each dependent variable are given in the following columns in two rows. The first row gives the estimate of the corresponding coefficient ($\hat{\beta}_i$), and the second row gives the $p$-value of the respective estimate. The adjusted coefficient of determination, $\bar{R}^2$, is given in the last column.

geographical and other resultant specifications on which the market performance also depends. It has already been shown in regression (13) that $s$ and $r$ can only have limited explanatory power. For most performance criteria, $\bar{R}^2$ is not even up to 30 % (see Table 4, the upper panel). Therefore, once after incorporating these specificities through other variables, such as the trading volume and the price volatility, a large proportion of the unexplained behavior has now been incorporated (see the significant increase in $\bar{R}^2$ from the lower panel of the same table). We find that after controlling the market characteristics the two capacities can indeed help enhance prediction accuracy. After incorporating the trading volume and the price volatility, $\beta_1$ and $\beta_2$ are both negative for all four accuracy criteria. In other words, conditional on the same trading volume and the price volatility, the higher the toleration capacity or the higher the exploration capacity, the better that the prediction market can predict.

## 4 Concluding Remarks

In this article, we address the issue of whether the better informed agent can help prediction markets in a spatial context. The better informed agents are characterized by their larger toleration capacity (sociability) and exploration capacity. The result is that under unconditional regression neither of them shows this enhancement, whereas, after controlling some market characteristics, the conditional regression shows their significance. Hence, in this sense, our paper shows that the quality of individuals does have a positive effect on information aggregation and on the formation of the wisdom of crowds.

The work can be extended in several directions. First, the network used here is a spatial network. In this digital age, given the significance of social groups in social media, it would be desirable to include a social network as part of the framework, and to study the effect of social network topologies. Second, the behavioral setting of the traders is very simple, i.e., the device of zero intelligence. It would be interesting to consider other behavioral settings involving cognition or learning, such as reinforcement learning or rule-based models. These extensions allow traders to base their decisions upon the information revealed in the order book. Third, the prediction market can be designed with other trading mechanisms, such as the call auction. It would be interesting to know whether these different trading mechanisms matter.

**Acknowledgments.** The authors are grateful to the two anonymous referee reports for their quite careful review of the papers. The paper has been revised by following many of the suggestions made in their painstakingly written reports. The remaining errors are, of course, solely the authors' responsibilities. The NSC grant NSC 101-2410-H-004-010-MY2 is also gratefully acknowledged.

## References

1. Chen, S.-H.: Varieties of agents in agent-based computational economics: a historical and an interdisciplinary perspective. J. Econ. Dyn. Control **36**(1), 1–25 (2012)
2. Chen, S.-H., Tung, C.-Y., Tai, C.-C., Chie, B.-T., Chou, T.-C., Wang, S.-G.: Prediction markets: a study on the Taiwan experience. In: Williams, L. (ed.) Prediction Markets: Theory and Applications, pp. 137–156. Routledge, London (2011). (Chapter 11)
3. Hayek, F.: The uses of knowledge in society. Am. Econ. Rev. **35**(4), 519–530 (1945)
4. Gode, D., Sunder, S.: Allocative efficiency of markets with zero intelligence traders: market as a partial substitute for individual rationality. J. Polit. Econ. **101**, 119–137 (1993)
5. Othman, A.: Zero-intelligence agents in prediction markets. In: Padgham, L., Parkes, D., Muller, J., Parsons, S. (eds.) Proceedings of 7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008), 12–16 May 2008, Estoril, Portugal, pp. 879–886 (2008)
6. Schelling, T.: Dynamic models of segregation. J. Math. Sociol. **1**, 143–186 (1971)
7. Smith, V.: Markets as economizers of information: experimental examination of the "Hayek Hypothesis". Econ. Inq. **20**(2), 165–179 (1982)

8. Tirole, J.: On the possibility of trade under rational expectations. Econometrica **50**, 1163–1182 (1982)
9. Yu, T., Chen, S.-H.: Agent-based model of the political election prediction market. In: Sabater, J., Sichman, J., Villatoror, D. (eds.) Proceedings of Twelfth International Workshop on Multi-agent-based Simulation, Taipei, Taiwan, 2 May 2011, pp. 117–128 (2011)