Learning with Prediction Markets: An Experimental Study

Cali Mortenson Ellis, University of Michigan Departments of Political Science and Public Policy, 5700 Haven Hall, 505 South State Street, Ann Arbor, MI, 48109-1045, cmortens@umich.edu
Rahul Sami, University of Michigan School of Information, North Quad, 105 S. State St, Ann Arbor, MI 48109, rsami@umich.edu

Abstract: Prediction markets are designed to aggregate the information of many individuals in order to forecast future events. These markets provide participants with an incentive to seek information and a forum for interaction, making them a promising tool to motivate student learning. We carried out a quasi-experiment in an introductory political science class to study the effect of prediction markets on student engagement. While we found no significant improvement in students' enthusiasm or extent of topical reading, we did find that those already reading broadly at the course start were more likely to trade actively in the markets. These findings indicate that prediction markets may be most successful as an education tool in settings where individuals are already knowledgeable about the topics of the market.

Introduction

Prediction markets (also known as information or decision markets) are designed to aggregate the information of many individuals in order to forecast future events. Such markets have been used in a variety of contexts; for example, the Iowa Electronic Market forecasts political and economic events, and the Hollywood Stock Exchange (*hsx.com*) forecasts movie box-office performance (Wolfers and Zitzewitz, 2004). The rationale behind using markets for forecasting is that traders reveal some information to each other through their trades, and learn from others' trading behavior, and thus, the market price reflects the combined information available to all traders. In this paper, we report on a quasi-experimental study of the use of prediction markets as a pedagogical tool to augment classroom learning.

Prediction markets have been used as classroom learning tools in keeping with a larger trend in education toward including interactive and technological resources into the classroom setting. Abramson (2010), for example, used prices posted on inTrade.com with upper-level undergraduates to illustrate American campaign strategies in the last Presidential election. Among the key advantages of prediction markets, researchers have noted that they provide incentives to motivate traders to "ferret out accurate information" and "not amplify individual errors, but eliminate them" (Sunstein, 2006). These strengths align well with our goals as instructors: we want to train our students to search for relevant information, and to critically analyze received information. Prediction markets also potentially provide a unique forum for cooperative learning, as they allow students to communicate with each other indirectly through market prices. Thus, students who may be unresponsive to face-to-face cooperative learning projects, such as classroom discussions or team assignments, may be still be able to collaborate and learn through the anonymous structure of the market interface.

The rest of this paper is structured as follows: In Section 2, we describe our experimental methodology. We explain the market structure and interface used in Section 3. We present descriptive statistics of our student population in Section 4. Section 5 reports on our data analysis procedures and the results of this analysis. In Section 6, we discuss the insights yielded by this analysis and the limitations of this study, as well as potential areas of further research.

Methodology

We employed a nonequivalent comparison group design, a quasi-experimental design using both control groups and pretests, as per Shadish, Cook and Campbell (2002). Students were randomized over the entire class into treatment (trader) and control (non-trader) groups. At the start of the semester, we administered a pre-test survey, pre-test knowledge quiz, and statement of informed consent to each student. We explained the purpose of the research and provided a live demonstration of the market interface using a simple example market. At the end of the semester, all students were also provided with a post-test survey and post-test knowledge quiz. Surveys and quizzes were administered in class, on paper in order to ensure a greater level of participation than can occur with web-based surveys. We also developed an extra credit question for one of the midterm exams. In order to ensure fairness and address potential concerns about diffusion, both treatment and control groups were given access to view the markets via the experimental website. However, only the treated group could trade in the market for cash prizes. We also relied on regular email communication with students to encourage them to trade in the markets. We report results for the 129 students for whom we collected complete information.

Market Topics

Prediction markets by definition are about current events with future uncertainty. We therefore created twelve markets with topics that fit the content of an introductory world politics course. We attempted to select a range of topics of interest to a variety of students, including whether Israel would announce a settlement freeze, if the elections in Haiti would take place as scheduled, and whether Google would cease all operations in China.

Market Structure and Interface

We used Zocalo, an open-source software package for running prediction markets, for our market deployment. We built a simplified custom user interface for these markets using the Drupal open source application platform. Students could see a news feed consisting of recent headlines relevant to the market; clicking on a headline led to the article on a newspaper website. We also included a place for students to leave comments and see comments by other students. Students could access the website at any time through university's secure portal.

Treated students received a budget of $1000 virtual dollars (equivalent to US$10 in real cash) at the start of the semester. For each of the market topics, there were two securities: a "yes" security that eventually paid off one virtual dollar if the event happened, and a "no" security that eventually paid if the event did not happen. Each market had a closing date which could be either the date of the supposed future event (for example, elections in Haiti which were scheduled to be held on February 28, 2010), a date in the middle of the term chosen by the researchers, or the end of the term. It was important to have some markets close in the middle of the term, as the students' correct predictions in such a market could generate additional points to be used for trading in another market. Students could use their point balances to buy or sell shares of either "yes" or "no" for the event in order to drive the price to a point that they thought was correct. For simplicity, we gave students the option of buying and selling shares in blocks of 10 or 100 units.

Students did not directly trade securities with each other; instead, an (automated) market maker bought securities from students who wanted to sell, and sold securities to students who wanted to buy. The market maker quotes a price at each point in time, and adjusts the prices for each unit bought or sold (Hanson, 2003).

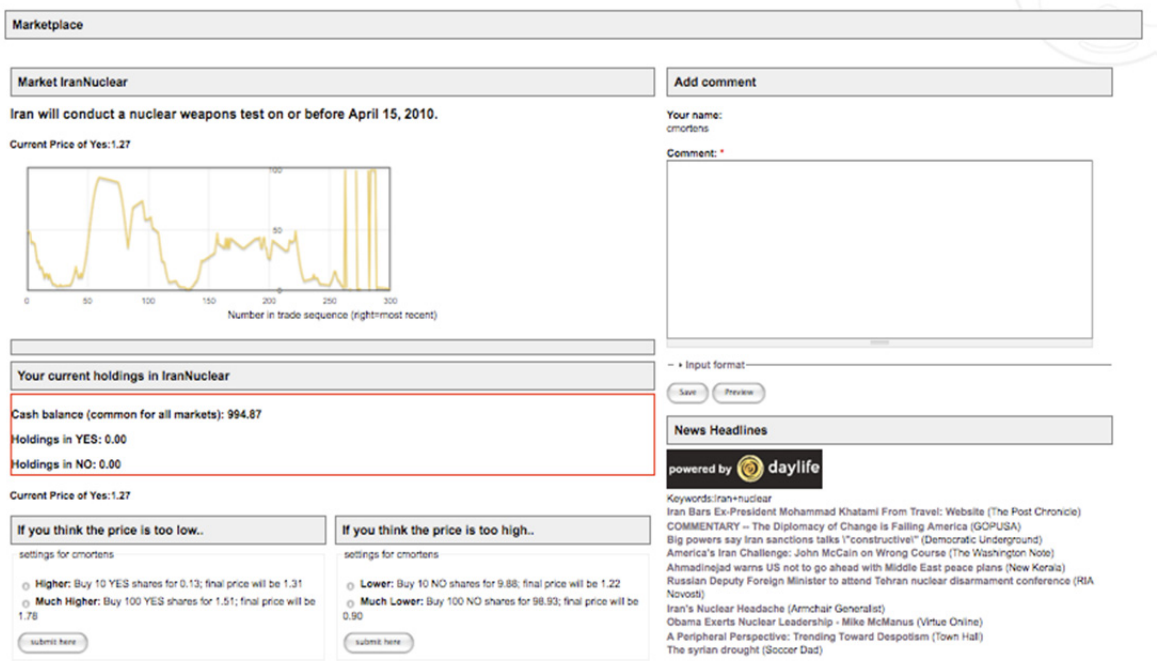

Figure 1. Screenshot of Market Interface.

Descriptive Statistics from Initial Knowledge Quiz and Survey

For this section on descriptive statistics, we report values for the overall population of the study, together with results of two-tailed t-tests with an assumption of unequal variances and the Satterthwaite approximation of degrees of freedom for mean differences between trader (treatment) and non-trader (control) groups. We use an α of 0.05 as the significance level for all of our statistical tests.

Survey Questions on Dependent Variables of Interest

Students filled out surveys that elicited information about their background, enthusiasm for the course topic, and sources they relied on for news. Our key dependent variables of interest pertained to enthusiasm, outside

reading, and general knowledge of world politics. Using a 7-point Likert scale to measure student's enthusiasm for world politics, we found an initial average score of 5.6202 *(|t|=0.076, p=0.94, d=0.01)*, corresponding to a point between "slightly enthusiastic" and "fairly enthusiastic." We asked if students read news about the politics of other countries at least once a week in a newspaper or on the internet. Almost 70 percent of students indicated that they did *(|t|=0.86, p=0.39, d=0.15)*, which may indicate a selection effect for students who elect this course in the first place. The quiz was the final component of interest, measuring actual knowledge about the market topics of the experiment. Examples of quiz questions include, "Iran has conducted a nuclear weapons test within the past year" and "The current Prime Minister of Japan promised the nation that he would soon visit The Yasukuni Shrine." The mean percent correct answers was 32.2 and there was no significant difference in mean performance between the control and treatment groups *(|t|=1.29, p=0.20, d=0.23)*.

MSLQ Survey Questions - Cognitive and Meta-cognitive Strategies

In order to control for the effect of intrinsic motivation, we administered a modified version of the Motivated Strategies for Learning Questionnaire (MSLQ, from Pintrich *et al*, 1993), specifically the section, "Cognitive and meta-cognitive strategies: self-regulation." This section includes questions such as, "When reading for a course, I make up questions to help focus my reading." These scores are demonstrably correlated with higher course grades (Pintrich *et al*, 1993). For each question, students are asked to rate on a 7-point Likert scale how much the statement applies to them. We hoped that higher scores on this section indicate students who may be more willing to take the extra steps necessary to ensure they thoroughly understand the subject matter. The mean score on the MLSQ was 3.684 *(|t|=0.39, p=0.69, d=-0.07)*.

Results

In this section, we present results on the effect of student exposure, obtained by comparing the initial and final surveys and knowledge quizzes of control and treatment group. Our original hypothesis was that participation in prediction markets would increase enthusiasm for and knowledge of the subject matter of an undergraduate political science course. Therefore, we expected to see improvements among traders in our primary dependent variables of interest: enthusiasm, outside reading and knowledge.

Table 1: Improvement in key dependent variables over one semester.

Question - Changes from start to end surveys	Non-Traders Start	Non-Traders End	Non-Traders Difference	Traders Start	Traders End	Traders Difference	*Diff-in-Diff (Trader-Nontrader)*
How interested are you in world politics? (1, completely unenthusiastic - 6, completely enthusiastic)	5.63	5.29	-0.34 *(|t|=1.80, p=0.077, d=-0.45)*	5.61	4.97	-0.64* *(|t|= 3.10, p=0.003, d=-0.76)*	-0.30 *(|t| = 0.079, p = 0.94, d=-0.01)*
Do you read news about the politics of other countries at least once a week in a newspaper or on the internet?	0.661	0.742	0.081 *(|t|=1.29, p=0.20, d=0.33)*	0.731	0.806	0.075 *(|t|=1.69, p=0.096, d=0.41)*	-0.006 *(|t| = 1.08, p= 0.28, d=-0.19)*
Quiz Scores (mean percent correct answers)	0.303	0.354	0.051* *(|t|= 2.19, p = 0.05, d=0.90)*	0.338	0.394	0.056 *(|t|=4.67, p= .056, d=0.79)*	-0.005 *(|t|=0.21, p=0.83, d=-0.04)*

Unfortunately, our results did not bear out our original hypothesis, as both groups of students reported lower levels of enthusiasm for the subject of the course, and students in the treatment group reported even less enthusiasm than the control group. We were not surprised to see a drop in enthusiasm given the timing of the end of semester survey shortly before the final exam. The first six columns of Table 1 report the results of a within-group paired t-test. Specifically, we test the null hypothesis that the difference between the measure from the beginning to the end of the semester is equivalent to zero. The last column of this table shows the differences between the change in means. This difference-in-difference between traders and non-traders was not significant for any of the key dependent variables, meaning that although some of the changes from the beginning to end of the semester were significant for each group, the changes in levels between traders and non-traders was not. Nonetheless, the greater drop in enthusiasm among traders is surprising and led us to seek a finer-grained perspective on the trader group.

Indicators of Active Trading

In this section, we report results only for students in the treatment group (traders), differentiating between *active* and *inactive* traders. Active traders (N = 45) are defined as those who logged in and executed at least one trade, as was required to receive the cash payment. Inactive traders (N = 22) are those in the treatment group who never logged in. Among the 42 students with multiple trades, the mean number of trades was 8.18. (We count multiple trading actions by a trader in the same market within one minute as a single trade.) In this section, we report differences between active and inactive traders in parentheses unless otherwise indicated.

Active traders had insignificantly higher MSLQ (3.77) scores than did inactive traders (3.40) *(|t|=1.50, p=0.14, d=0.39)*, and their score was comparable to the mean for the control group (3.72). Recall that the MSLQ locates students who are more likely to, "use more deep-processing strategies such as elaboration and organization and who attempt to control their cognition and behavior through the use of metacognitive planning, monitoring, and regulating strategies" (Duncan and MacKeachie, 2005). It is possible that participation in the prediction markets was seen as an additional metacognitive strategy used to supplement course preparation.

Another notable finding among active traders is that they had a significantly higher prior level of reading in the area. At the start of the semester, 82 percent of those students who would eventually become active traders reported reading about the politics of other countries at least once per week. By comparison, only 54 percent of inactive traders reported such reading *(|t|=2.25, p=0.031, d=0.83)*. In other words, the students who were already reading broadly at the start of the course were more likely to trade actively in the market. By the end of the semester, only inactive traders exhibited a statistically significant improvement of 18 percent on this measure. Active traders improved their reading by 7 percent, which is significant at the $\alpha = 0.1$ level, and remained ahead of other groups, with 84 percent reporting this activity by the end of the term. There was no statistically significant difference in improvement between active and inactive traders or traders and non-traders.

Table 2: Reading about the politics of other countries at least once per week (active vs. inactive traders).

	Semester Start	Semester End	Difference (Start-End) within-group t-test
Active Traders	0.822	0.844	0.0222 *(|t| = 0.443, p = 0.66, d=0.13)*
Inactive Traders	0.545	0.727	0.1818* *(|t| = 2.16 p = 0.042, d=0.92)*
Difference (Active-Inactive) independent 2-group t-test with unequal variance	*0.277* (|t|=2.25 p=0.031, d=0.83)*	*0.117 (|t|=1.05 p=0.30, d=0.37)*	*-0.1596 (|t| = 1.63, p=0.11, d=-0.57)*

Table 3: Improvement in knowledge quiz scores by group (active vs. inactive traders).

	Start Quiz Score	End Quiz Score	Difference (Start-End) within-group t-test
Active Traders	0.329	0.387	.0572* *(|t| = 2.67, p = 0.011, d=0.79)*
Inactive Traders	0.354	0.407	.0532 *(|t| = 1.75, p = 0.095, d=0.74)*
Difference (Active – Inactive) independent 2-group t-test with unequal variance	*-0.024*	*-0.020*	*-.00397 (|t| = 0.107, p = 0.92, d=0.036)*

We next looked at the performance on the quiz score (Table 3). Active traders showed slightly greater progress on this measure, increasing from 32.96% to 38.68%, a statistically significant 5.72% improvement. These results taken together indicate that the prediction markets may have some value in improving content learning for students who are motivated by the instrument, but further experiments are needed to investigate the extent to which they enable learning beyond the specific topics of the markets they trade in.

Motivations and Behavior of Active Traders

The following section only reports findings for students in the group of active traders (N=45), defined as those who logged in and executed one or more trades. We wanted to know if students discussed prediction markets or the topics of the markets outside class, an indirect indicator of collaborative learning. 31% of traders reported discussing the market topics with friends or family and 24% reported discussing the prediction markets experiment itself with other classmates. 80% found the experiment relevant to the topic of the course.

We asked active traders if the experiment encouraged them to read about the *market topics* outside of class. 51% reported reading more about the Google/China dispute, and 40% reported reading more about the Olympics. Only 11% reported reading about the military or Palestinian elections, and 9% reported reading about the elections in Haiti. Overall, there was some evidence that the markets prompted additional reading, but that it was not distributed evenly across all the topics. This is consistent with the original design of our experiment, where we attempted to provide markets on a range of topics to accommodate different interests.

We also asked students how they made their trading decisions. Fifty-eight percent stated that trading decisions were made "based on personal beliefs," followed by 51% based on news reports. The smallest number of students decided based on the outcome they wanted (4 percent) or the trades of others (i.e. the price reported on the graph – 6 percent). Forty percent of students believed that their predictions were more correct than those of their classmates. This is in line with the claim by Sunstein (2006) that market trading encourages individual information processing and reporting, rather than groupthink or herd behavior.

Conclusion

As Whitton (2010) warns, not all students will be responsive to interactive technological projects such a prediction markets. One of our most striking findings is that active traders had a significantly higher prior level of reading than other students in the topics of the markets, outside of their course requirements, both at the beginning and end of the semesters. Another is their higher level of MLSQ characteristics, indicating students who have more self-regulated study habits. The students who did actively trade reported that they enjoyed doing so, and that it prompted them to read more widely on some of the market topics. These findings together hint that prediction markets may be best deployed in a classroom of students who are highly motivated and already engaged in the subject matter. An elective upper-level undergraduate course or a graduate course may be more appropriate settings for using prediction markets as an educational tool. Our experiment also provides some guidance on the design of a prediction market interface: The feedback we received from traders also suggests that it is useful to integrate news feeds with the market.

For reasons of fairness in our randomized study, we chose a controlled set of market topics that were outside of the specific topics studied in the class. One unanswered question from our research involves how students would interact with a decision market which is more closely connected to their interests. For example, it is possible to design the interface so that students can create markets based on their own interests in certain topics of a class. We asked students directly if they had the opportunity to create their own market on any topic (for example, who will win the NCAA championships), if they would do so, and 44 percent stated that they would. It would also be interesting to conduct a future study with markets linked to the course syllabus, for verification of whether more immediacy in the market topics would pay off in learning gains.

References

Abramson, P. (2010). Using Intrade.com to Teach Campaign Strategies in the 2008 U.S. Presidential Election. *PS: Political Science & Politics*, 43:139-144.

Duncan, T. G. & McKeachie, W.J. (2005). The Making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, 40(2): 117-128.

Hanson, R. (2003) Combinatorial Information Market Design. *Information Systems Frontiers*, 5(1): 107–119.

Pintrich, P., Smith, D., Garcia, T., & McKeachie, W. (1993). Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement,* 53(3):801-813.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin.

Sunstein, C. R. (2006). Deliberating Groups versus Prediction Markets (or Hayek's Challenge to Habermas). *Episteme: A Journal of Social Epistemology,* 3.3: 192-213.

Whitton, N. (2007). Motivation and computer game based learning. In *ICT: Providing choices for learners and learning. Proceedings ascilite Singapore 2007*. Available at http://www.ascilite.org.au/conferences/ singapore07/procs/whitton.pdf

Wolfers, J. & Zitzewitz, E. (2004). Prediction Markets. *Journal of Economic Perspectives*, 18(2): 107-126.

Acknowledgements

This research was partially supported by the National Science Foundation under grant CCF-0728768. We are very grateful to Allan Stam for supporting these experiments in his courses; Skip Lupia, Stephanie Teasley, Kara Makara, and Rick Wilson for numerous helpful suggestions and feedback; Stan Dimitrov for assistance with the market interface; and Mohammed Hadhrawi, Yung-Ju Chang, and Yi-Chin Wu for useful discussions on an earlier project. This work was examined by the Institutional Review Board of the University of Michigan.