Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence

Instructor Rating Markets*

Mithun Chakraborty and **Sanmay Das** and **Allen Lavoie**
{chakrm,sanmay,allenbl}@cs.vt.edu
Virginia Tech

Malik Magdon-Ismail
magdon@cs.rpi.edu
Rensselaer Polytechnic Institute

Yonatan Naamad
ynaamad@cs.princeton.edu
Princeton University

Abstract

We describe the design of *Instructor Rating Markets* (IRMs) where human participants interact through intelligent automated market-makers in order to provide dynamic collective feedback to instructors on the progress of their classes. The markets are among the first to enable the empirical study of prediction markets where traders can affect the very outcomes they are trading on. More than 200 students across the Rensselaer campus participated in markets for ten classes in the Fall 2010 semester. In this paper, we describe how we designed these markets in order to elicit useful information, and analyze data from the deployment. We show that market prices convey useful information on *future* instructor ratings and contain significantly more information than do past ratings. The bulk of useful information contained in the price of a particular class is provided by students who are in that class, showing that the markets are serving to disseminate insider information. At the same time, we find little evidence of attempted manipulation by raters. The markets are also a laboratory for comparing different market designs and the resulting price dynamics, and we show how they can be used to compare market making algorithms.

Introduction

Prediction markets are an example of a venue where humans interact with artificial agents (like market-makers) in order to form a collective intelligence. The humans in the market may have private information, while artificial agents serve to lubricate the functioning of the market by incentivizing the revelation of that information. This paper presents a novel application of prediction markets to instructor evaluations. Such markets have the potential to provide *dynamic* feedback on the progress of a class. While the instructor is only rated occasionally, price movements can provide continuous feedback, in the same way that prices in election prediction markets provide feedback to a campaign on its successes and failures.

Typical large prediction markets, such as election markets, attempt to predict a stable statistical aggregate quantity: voting turnouts range from the tens of thousands to the tens or hundreds of millions. In contrast, a class may have somewhere from 5-100 raters, so any individual in the class simultaneously has a lot more information and can affect the outcome much more substantially than in traditional large markets. This raises two questions: (1) Will the information get disseminated effectively? (2) Will students try to manipulate the ratings or the markets so that we can no longer trust the information in either the markets or the ratings? We need experience with medium-scale prediction market deployments like the IRMs in order to begin to address these questions. In addition, we can also use the IRMs to study the effects of market design on information revelation.

In this paper, we describe a pilot deployment of these markets at Rensselaer Polytechnic Institute in the Fall of 2010, with more than 200 students participating across 10 classes. We answer Question (1) above in the affirmative: we find that prices are, in fact, predictive of *future* instructor ratings, and significantly more predictive than are previous ratings, showing that they incorporate *new* information. The higher predictivity is due to the trades of insiders: when previous and future liquidations differ, students who are enrolled in a class trade in the direction of future liquidations while others trade in the direction of the last liquidation. We also provide evidence related to Question (2): at least in our context, there is little evidence for manipulation: prices predict ratings in the IRMs, and the IRM ratings turn out to be very well-correlated with the independent "official" university ratings of instructors. Finally, we show how to use the IRMs to study market design: we compare a Bayesian market-making algorithm (BMM) with the standard Logarithmic Market Scoring Rule (LMSR) and show that BMM can provide more price stability than LMSR while also making more profit.

Related Work. In recent years, prediction markets have gone from minor novelties to serious platforms that can impact policy and decision-making (Wolfers and Zitzewitz 2004a). There has been a concomitant rise in interest in prediction markets across academia, policy makers, and the private sector (Wolfers and Zitzewitz 2004b; Berg and Rietz 2003; Servan-Schreiber et al. 2004; Arrow and others 2007; Chen and Pennock 2007). There has been some research on the design and deployment of live prediction markets (Berg et al. 2008; Othman and Sandholm 2010; Cowgill, Wolfers, and Zitzewitz 2010). There have been small ex-

*An abstract appeared in the Proceedings of the Second Conference on Auctions, Market Mechanisms, and Their Applications. Markets run while the authors were at Rensselaer.

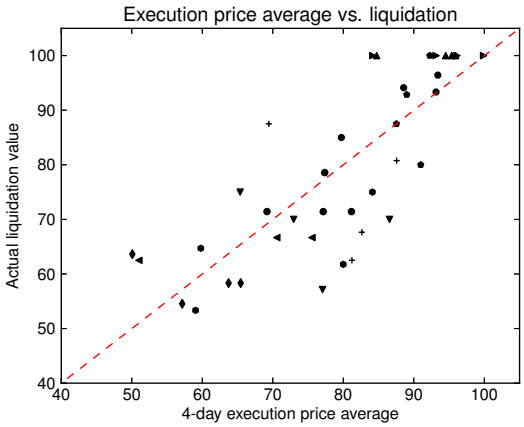

Figure 1: Traded prices predict liquidation values well.

periments to test the impact of insiders on small, short-running, experimental prediction markets (Hanson 2008; Hanson and Oprea 2009). The IRM is more of a "field experiment" than these controlled studies. It is significantly longer in duration and larger in the number of participants. It is also unique in its goal of explicitly providing dynamic feedback to instructors that can be correlated with real measures of performance.

A second motivation of this work is to provide a framework for comparing prediction market structures. There has been little systematic work in this area. While much of the literature on liquidity provision discusses the pitfalls and advantages of different algorithms (Pennock and Sami 2007; Othman et al. 2010; Chen et al. 2009; Pennock 2004), only recently have there been attempts to simultaneously compare market microstructures in controlled experimental designs involving human traders (such as the work of Brahma et al. (2012)). However, Brahma et al. make these comparisons using short, ten-minute experiments. We open the door to studies of such issues in longer-horizon markets.

Description of the Markets

We ran virtual cash markets for 10 different classes during the fall semester of 2010. The experiment was divided into five periods of approximately two weeks each, with one period extended due to a holiday break. Each instructor-course pair was one security (and a fresh security was instantiated for each instructor-course pair at the beginning of each of the five periods). Each security could be traded by anyone at the institute. At the end of each period, students enrolled in a particular class rated their instructor. The payoff (liquidating dividend) of the security for that instructor-course pair for that period was the average rating given by the students.

Ratings. Students taking one of the ten subject classes were given keys at the beginning of the semester which enabled them to rate their instructor at the end of each trading period. Each student who registered a key was sent a reminder email for each trading period. For the first period, rating was done through the same website as trading; for the remaining four periods, students could also rate directly from their reminder email. Initially, rating could be either thumbs-up (100%) or thumbs-down (0%), but a neutral option (50%) was added beginning with the third period. The initial limitation reflected the idea that only 0 and 100 were rational choices for traders seeking to maximize their wealth; we relaxed this limitation in response to feedback from students who did not want to rate their instructor either positive or negative. The liquidation value $\in [0, 100]$ of a market was the average of all ratings cast for the associated trading period.

Incentives. After each liquidation at the end of a trading period, a trader's account value was equal to their cash balance plus the liquidation value of any shares they held. All trader accounts were then re-initialized for the next trading period (there was no carryover).

Prizes were awarded twice: once after the second period of trading, and once after the fifth period of trading. Six prizes were raffled off each time, based on a trader's rank and account value in each period. The top 3, 5, 10, and 20 accounts in each period were eligible for the 1st, 2nd, 3rd and 4th prizes respectively.

Periods	Prizes					
	1st	2nd	3rd	4th	5th	6th
1-2	$69	$49	$40	$30	$20	$20
3-5	$150	$100	$60	$40	$20	$20

Table 1: The value of awarded prizes.

If an account featured in the top 3 accounts of a period, it was eligible for all the prizes; if an account featured in the top 5 (but not top 3), it was eligible for all prizes but the top prize, etc. The fifth prize was a participation prize awarded uniformly at random to one of the top 50% of traders in each period. The sixth prize encouraged participants to rate their professors and was drawn with probability proportional to the number of times a trader provided ratings. The prizes were awarded from 1st to 6th, with the restriction that once an account was awarded a prize, it became ineligible for any subsequent prize. Prize values are summarized in Table 1.

From a theoretical standpoint, these incentives create complex utility functions. We could instead have awarded prizes with probability proportional to a trader's total account value. However, making such a scheme sufficiently rewarding was not practical given reasonable constraints on the value of awarded prizes; rank order incentives such as those used in the IRM can be significantly more effective than proportional payments (Luckner and Weinhardt 2007) due to decreased risk aversion in traders. Simply paying participants based on their performance without vastly increasing the total amount awarded would likely have been demotivating (Gneezy and Rustichini 2000). While linear rewards for participation would seem to at least yield simple incentives, even this does not occur in practice, as other motivations are found to have a significant impact on experiment participants (Loewenstein 1999).

Microstructure. Traders interacted with the markets by placing market orders through a Web interface. Traders were presented with a full history of traded prices and liquidation values for each security, along with links to the associated course website. They were also shown the current (spot)

price of the security, and could place a market buy or sell order for a desired quantity – they would then receive a price quote for their entire order, and were asked to confirm. For the first two periods, users started with 50000 units of virtual currency and 50 shares of each market. For the final three periods, users started with 100 shares of each class and the same amount of currency.

Price quotes were generated using two different market making algorithms (only one algorithm was used for any given market during a particular trading period). We used an implementation of Hanson's logarithmic market scoring rule (LMSR) (Hanson 2007) with a b parameter of 125 (restricting loss to 8664.34 in any given period)[1], and an implementation of the Bayesian Market Maker (BMM) described by Brahma et al. (2012). Both market makers are initialized at the beginning of a trading period so that the quoted price in each market is the same as the close of the previous trading period.

Market Participation. Overall there were 226 registered users, with registration limited to current RPI students, faculty, and staff. Of these, 198 users made at least one trade during the experiment. Participation declined as the experiment progressed, with 117 active traders in the first period and only 33 in the fifth period. Rating was more steady, peaking at 93 raters during the second period, but never dropping below 70 raters during any period. The backgrounds of participants were mixed: from undergraduates studying physics to faculty in computer science.

Information Content of Prices

Prediction markets attempt to aggregate information and to incentivize the dissemination of information that is otherwise difficult to obtain. One question is whether traded prices in the IRM provide any new information about *future* instructor ratings. If traders simply provide a noisy realization of the previous rating (dividend), for example, then the prices themselves do not provide useful dynamic instructor feedback. Do the markets have predictive power?

Predictivity of prices. Figure 1 answers this question in the affirmative. In the figure is a scatter plot of the upcoming liquidation value versus the average traded price. Different markets are referenced with different symbols. Also shown is the ideal outcome (the line $y = x$). Modulo noise in the data, there is good agreement between the data and the ideal line. We use an average traded price because prices are noisy and averaging can provide a better proxy for the market value than the price of any single executed trade.[2] Such smoothed prices were significantly more predictive than previous liquidation values, with a four-day average price yielding an R^2 value of 0.58, while previous liquidations produced an R^2 of 0.48. This finding is robust to different averaging windows for prices and different aggregates for previous liquidations.

To further validate that market prices are a better predictor of future liquidation values than prior prices, we ran a regression using both the previous liquidation and the market price as independent variables in the following model:

$$\mathrm{Liq}_{s,\rho} = \beta_1 \mathrm{Liq}_{s,\rho-1} + \beta_2 \mathrm{Price}_{s,\rho} + \alpha \qquad (1)$$

where $\mathrm{Liq}_{s,\rho}$ is the liquidation value of market s in period ρ, and $\mathrm{Price}_{s,\rho}$ is the 4-day average market price before liquidation. The sample size for this regression is 40, since we have no previous liquidation value for securities in the first period.

The significance of the previous liquidation value at the $p = 0.05$ level disappears when price is included in the linear model above, showing that previous liquidation value provides no additional information beyond price in this regression. This result is robust with respect to the choice of how price is smoothed. For $\mathrm{Price}_{s,\rho}$ equal to the 4-day average price, we find that β_2 is the only statistically significant coefficient (at $p < 0.05$). The results are qualitatively unchanged when adding random effects controls for per-period and per-stock variations.[3]

Insider trading/sources of information. Having shown that prices are predictive, we would like to know where the new information is coming from. While this is sometimes done by looking at the trade prices of different types of traders, that methodology is more appropriate for markets with limit orders. In a market-maker mediated market, it makes more sense to look at the directions of trades. Consider a single trade on the IRM: either this trade moves a price toward the corresponding instructor's future rating, or away from it. By examining the set of all IRM trades in this manner, we can get an idea of the information revealed by groups of traders. We would expect that in-class traders, since they determine instructor ratings, would provide more information than out-of-class traders. Indeed, in-class traders traded toward the future liquidation 53.9% of the time (95% confidence interval 53.0% to 54.8%), while out-of-class traders traded toward the future liquidation only 52.5% of the time (95% confidence interval 52.3% to 52.8%). The difference is statistically significant ($p = 0.015$). This tells us that in-class traders brought more information to the IRM. However, we know that previous liquidations are a good predictor of future liquidations; how many of these trades are simply based on old information?

To determine which traders bring *new* information to the IRM, we can examine trades that occur at prices between the previous liquidation price and the future liquidation price (see Figure 3). In such situations, if insiders are truly the sources of fresh information, we would expect them to trade more in the direction of the future liquidation, while others trade more in the direction of the last liquidation. Examining the data confirms this hypothesis. In situations where

[1] From a market making perspective, a real-valued dividend $\in [0, 100]$ is equivalent to the more typical 0-1 dividend, modulo a constant factor; the price computation for LMSR is exactly the same, and the loss bound is determined by the extreme values of the dividend.

[2] Collecting information from prices in this pilot deployment would have been difficult to do in real time because of volatility, especially when using LMSR as the market maker. One could combat this by increasing the loss tolerance of LMSR, effectively performing smoothing with the market maker itself.

[3] We added α_s and α_ρ as random effects, assumed to be normally distributed with mean zero, representing random per-stock and per-period variations respectively.

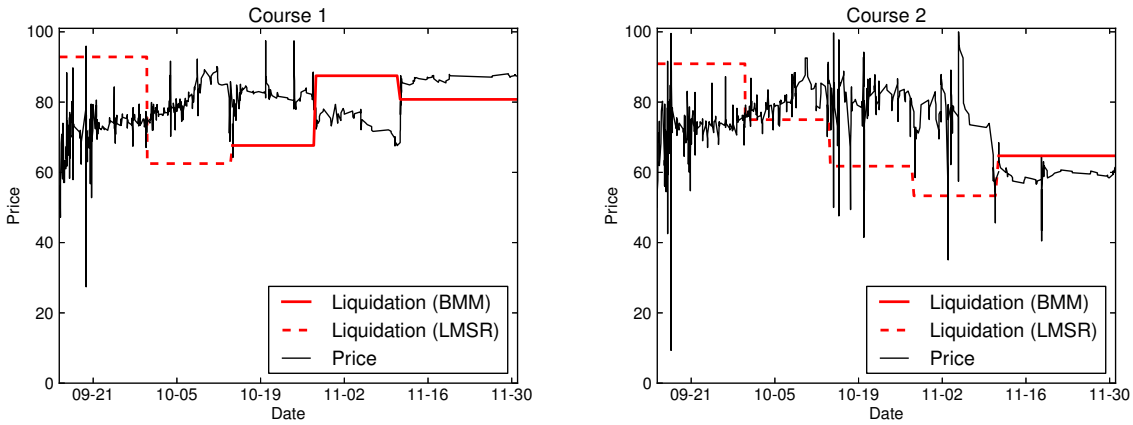

Figure 2: Price charts and liquidation values for selected markets, with line style indicating the market making algorithm. Each trade is plotted according to its transacted price with no smoothing.

the execution price was in between the last liquidation value and the next liquidation value, in-class traders traded toward the future liquidation 53.5% of the time (95% confidence interval 51.7% to 55.2%, so also significantly more than 50% of the time). Out-of-class traders favored the previous liquidation, trading toward the future liquidation only 47.7% of the time (95% confidence interval 47.0% to 48.4%, so significantly less than 50% of the time). The difference is, of course, statistically significant ($p = 7.1 \times 10^{-7}$). This is compelling evidence that out-of-class traders were mostly trading on old information, and the markets serve to disseminate the inside information of in-class traders to the world, and provide feedback to instructors in doing so.

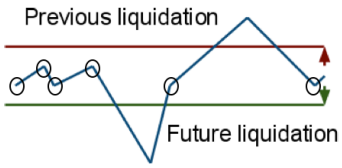

Figure 3: Methodology for determining who brings new information to the market. Trades that occur between the previous and future liquidation prices are circled, and move the price either in the direction of the future liquidation (new information) or the past liquidation (old information).

Qualitative features of prices. Figure 2 shows the traded prices and liquidation values for a selection of markets (traders saw this information in a similar format, although they were not aware of which market maker was used in which period). The figures highlight certain interesting qualitative features of the price processes. First is the effect of volatility, which may make the instantaneous price a less useful piece of information for the instructor at any point in time than a smoothed version of the price, as discussed earlier. An alternative would be to use a less volatile market making scheme (different parameters or a different algorithm). In fact, volatility does appear to be significantly less for markets using BMM (Brahma et al. 2012), which we discuss later; this lower volatility does not come with any loss in predictive ability of the resulting prices. Second, prices often move towards the previous liquidation right after that value is revealed, without moving all the way there. We see this behavior clearly in Course 1, especially during periods four and five. Two of the IRM classes always liquidated at a value of 100, and in these classes the security prices slowly converged to 100; the slow rate of convergence is probably because the incentive to buy a security near 100 even given a sure liquidation at 100 is very small.

Summarizing the evidence: the markets are useful and predictive, providing information on future ratings that instructors will receive. We find strong evidence that most of the useful new information is added by in-class traders. Meanwhile it appears that out-of-class traders help in providing market stability by trading toward previous liquidation values, offsetting large noise trades.

Trading and Rating Behavior

One of the unique benefits of the IRM is that we have data on both the trading and rating behavior of the participants. This allows us to explore issues in market manipulation and trader behavior in ways that were previously not possible. For example, we present evidence not only that the IRM succeeded in its primary goal of providing dynamic predictive information on how a professor is doing, but also that this information was mostly provided by students enrolled in the class. Here we look more deeply into the behavior of users.

Insiders, Manipulation, and Collusion. Traders who had rating credentials in a market (in-class traders, or insiders) could both trade in the market and affect the dividend through their rating. Therefore, not only did they have better information on the professor being traded than other participants, but they had the opportunity to explicitly choose how to rate the professor based on their position in the stock. We define "manipulation" as situations in which students provide a rating they do not truly believe in order to maximize their profits from the IRM. There were plenty of opportunities for manipulation: several classes had only 3-5 raters, and information on how many ratings contributed to a par-

ticular liquidation value was made easily accessible on the trading interface (along with the prior liquidation values), allowing raters to estimate their impact on a market's liquidation. Of course, knowing if manipulation actually occurred is difficult, but we provide several pieces of data that make the case that there was little manipulation.

IRM Ratings Were Not Manipulated. Do IRM student ratings correspond well with what they actually thought of the class? Since seven of the ten classes were in the Computer Science Department, we were able to measure the correlation of IRM ratings with the official end-of-semester student evaluations.[4] We averaged the ratings and prices of periods 3-5 in the IRMS. The coefficient of correlation of the IRM ratings to the official ratings for these 7 classes was 0.86, and the coefficient of correlation of the IRM prices averaged over these periods with the official ratings was 0.75. The strong correlation between IRM ratings and official ratings validates the usefulness of our markets in terms of a real benchmark that is "outside the system," and also indicates that students were rating honestly in the IRMs, and that we do not need to worry about experiment-wide misbehavior.

Little Evidence For Manipulation in IRM Prices. We considered any group of raters who both gave the same rating and made a significant amount of money (1000 virtual currency each) trading the associated security during a given period as candidates for having colluded. We observed collusive behavior in course 2 during period 4. A group of 3 raters together made about 9000 in virtual currency by selling course 2's security and rating the course low. These 3 students controlled 20% of the liquidation value; since most liquidations were between 60 and 100, this was enough for the manipulators to reduce the security's price significantly below the market's expectation. This liquidation was Course 2's lowest, although it is not apparent from the liquidations alone that manipulation was involved (see Figure 2). Pairs of raters made somewhat smaller amounts of virtual currency in several other markets, but it is not clear if intentional manipulation was involved.

More surprising than the observed manipulation in the IRM was its relative scarcity. Most markets did not see any successful collusion based on the criteria that raters both made money and rated together during a given period. Perhaps students did not understand the opportunities for manipulation, or perhaps giving accurate feedback was more important than winning prizes for some raters.

We note that the potential for manipulation was not limited to groups or to simple rating manipulation. Examining the trading records of raters who made more than 1000 virtual currency trading in a given security during a given period, however, seems to indicate that such opportunities where not successfully exploited; we do not observe significant shifts in trading activity by these raters. Manipulation by non-raters seems significantly less likely given the relative lack of information and influence.

Trading Strategies and Profits. Traders varied wildly in their activity levels, strategies, and apparent rationality. While some amassed large quantities of virtual currency by frequently monitoring for mispriced securities, others seemed eager to cause as much havoc as possible while divesting themselves of their entire initial capital. Figure 4 shows the number of trades and the number of shares traded per user and per period, grouped by the user's account value at the end of that period. We see that the defining feature of the most successful traders was activity; while they did trade more shares overall, they did so in almost twice as many transactions as the less successful traders. The worst traders also stood out, making a moderate number of massive trades.

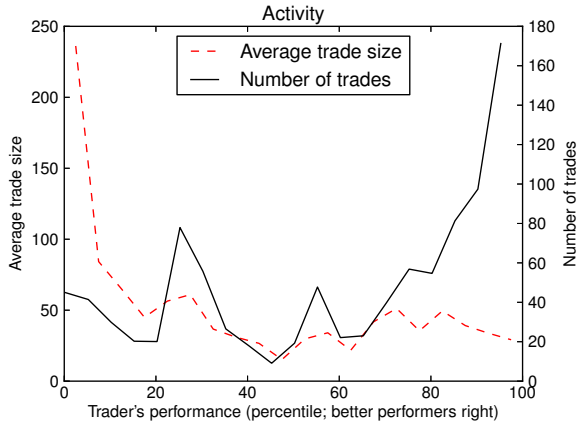

Figure 4: Successful traders made many smaller trades.

Effects of Microstructure

The IRM is a powerful platform for testing the effects of different microstructures on price dynamics. We tested two different market-making algorithms. Brahma et al. (2012) develop a Bayesian Market Maker (BMM) (building on (Das and Magdon-Ismail 2009; Das 2008; 2005)), and compare with Hanson's Logarithmic Scoring Rule (LMSR) market maker (Hanson 2007). In short, 10 minute trading sessions, they find that BMM can offer comparatively higher price stability and smaller spreads than LMSR without suffering losses in expectation. On the flip side, LMSR comes with a strong loss bound, while BMM may occasionally take high losses. We provide additional evidence for these conclusions in a more realistic long-running experiment, with ample opportunities for strategizing and manipulation.

Description of LMSR and BMM. LMSR is a purely inventory-based market maker. For a single security with payoff in $[0, 1]$ (as noted above, the cost/price function and loss-bound is exactly equivalent to the case of binary payoffs), the spot price at an inventory level q_t is given by $p(q_t) = e^{q_t/b}/(1 + e^{q_t/b})$, where b is a positive parameter, and the cost for a change Q in the inventory is $C(Q; q_t) = b \ln \left[(1 + e^{(q_t+Q)/b})/(1 + e^{q_t/b})\right]$. Thus, for a buy or sell order of size Q at an inventory level q_t, the market maker quotes a volume weighted average price (VWAP) $|C(Q; q_t)/Q|$ where Q is positive for buys and negative for sells. The inventory is updated to $(q_t + Q)$ only if the trade is accepted, and the market maker waits for the next order.

[4] To protect instructor confidentiality, we gave the IRM ratings and prices to the Department Head, who ran the correlations against end-of-semester student evaluations.

	Periods	Avg profit	Max loss	Std	Liq dev
LMSR	35	1341.67	-5298.58	8.6	16.9
BMM	15	8273.13	-13763.40	3.0	9.6

Table 2: Overview of statistics for LMSR and BMM, showing average profit, max loss, the standard deviation of prices, and deviation of prices from the market's liquidation value.

Note that, in our implementation, all these quantities are multiplied by 100 to keep the prices in the range $[0, 100]$.

BMM, an information-based market maker, maintains a Gaussian belief distribution $N(\mu_t, \sigma_t^2)$ for the value of the market; the spot price is equal to the mean belief μ_t. The underlying assumption is that trader valuations are normally distributed around the true value V. A fixed trade size parameter (α) determines quoted prices: every buy/sell order of size Q is imagined to be a sequence of $k = \lceil Q/\alpha \rceil$ independent mini-orders of sizes $\{\alpha_i\}_{i=1}^{k}$ which are all α except possibly the last one. The market maker then quotes a VWAP and updates its state depending on the trader's decision (acceptance/cancellation); the precise updates are non-trivial, but efficient (see (Brahma et al. 2012) for details). Even though the Bayesian belief updates converge, BMM can adapt to market shocks, where the market's value changes dramatically. To do so, BMM maintains a "consistency index" that quantifies how consistent the trades in a window of size W are with the current belief. When trades are inconsistent with the belief, the belief variance rapidly increases, allowing quick adaptation.

LMSR is simple and loss bounded: the loss is at most $b \ln 2$. Moreover, being inventory-based, it is difficult to manipulate; and, assuming rational traders who learn consistently from prices, an LMSR-mediated market converges to a rational expectations equilibrium. Though the loss is bounded, LMSR does typically run at non-zero loss. One drawback is that a single parameter b controls various aspects of the market such as the loss-bound, liquidity, and adaptivity; therefore, achieving a trade-off can be difficult. Moreover, Brahma et al. find (and we confirm here) that if the beliefs of the trading population do not converge, prices can be very unstable. BMM, on the other hand, is not loss-bounded but makes much less loss in expectation while providing an equally liquid market. Moreover, in the absence of market shocks, BMM's belief (and hence the spot price) converges owing to a monotonically decreasing variance, even if the traders maintain heterogeneous valuations.

Exploiting BMM. The variance of BMM's belief distribution determines its spread. A simple implementation can be manipulated by artificially tightening the spread, with a sequence of alternating small orders followed by a large order to exploit the low spread. To avoid this, we perform inference on BMM's variance parameter only once for each trader unless an intervening trader also places an order. This idea can be easily extended to pairs of colluding traders, but could suffer from Sybil attacks. Such manipulation strategies are highly non-obvious, and, further, we limit traders to a single account by requiring an institute email address for authentication. Ultimately, exploitation of BMM did not become an issue.

Comparison of Market Makers. We confirm the major findings of Brahma et al.'s previous comparison of BMM to LMSR. In essence, BMM offers more stable prices (see Figure 2 and Table 2), while making higher profits and maintaining lower spreads (see Table 2). We set LMSR's b parameter to 125; by increasing b one can get lower spreads and more stability, but at the expense of other tradeoffs. For example, the b parameter of LMSR is an explicit market subsidy, increasing not only the loss bound but the expected loss of the market maker in reaching a given equilibrium price. Since LMSR actually *made* money on average[5], this could be an acceptable tradeoff. BMM already made more money on average in the IRM, however, and so comparing volatility is quite reasonable. It is interesting to note that the median trader made money when trading with LMSR, although the mean was below 0, whereas both the mean and median traders lost money with BMM. The volatility of prices and the deviation from the future liquidation value suggest that not only was the BMM price more stable than that of LMSR, it also provided a better estimate of the liquidation value. These results are robust and significant when regressing with per-security random effects.

Discussion

The Instructor Rating Markets are a field experiment in the space of agent-mediated prediction markets that incentivize humans to truthfully reveal their information, and, in doing so, provide useful dynamic feedback (in this case to instructors). The IRMs are a platform for studying the behavior of insiders and potentially manipulative participants in unprecedented depth. Many of the questions we study here would not be amenable to either short, intensely controlled lab experiments, or to study based on the data from prediction markets deployed "in the wild." Perhaps the most fruitful questions to pursue in similar, medium-sized experiments in the future revolve around manipulation and the role of market design (including the design of automated market mediators) in achieving good information dissemination properties.

Acknowledgments. We are particularly grateful to Professors Cutler, Kar, Newberg, Goldberg, Varela & Wallace for their participation in the IRM, and Prof. Hardwick for his support of the project and for running correlations with official evaluations. We also thank our anonymous traders!

This research is supported by an NSF CAREER award (IIS-0952918/1303350) to Sanmay Das. Malik Magdon-Ismail is partially supported by U.S. DHS through ONR grant N00014-07-1-0150 to Rutgers. The content of this paper does not necessarily reflect the position or policy of the U.S. Government, no official endorsement should be inferred or implied. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

[5]This does not contradict our previous results concerning the information content of prices. While LMSR only makes money when the first market price is more informative than the last, we find that an average of the last few prices is more informative still. Setting the last price to this average and computing the hypothetical profit, LMSR loses money on average and in most cases.

References

Arrow, K., et al. 2007. Statement on prediction markets. AEI-Brookings Publication No. 07-11.

Berg, J., and Rietz, T. 2003. Prediction markets as decision support systems. *Information Systems Frontiers* 5(1):79–93.

Berg, J.; Forsythe, R.; Nelson, F.; and Rietz, T. 2008. Results from a dozen years of election futures markets research. *Handbook of Experimental Economics Results* 1:742–751.

Brahma, A.; Chakraborty, M.; Das, S.; Lavoie, A.; and Magdon-Ismail, M. 2012. A Bayesian market maker. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, 215–232.

Chen, Y., and Pennock, D. 2007. A utility framework for bounded-loss market makers. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, 49–56.

Chen, Y.; Dimitrov, S.; Sami, R.; Reeves, D.; Pennock, D.; Hanson, R.; Fortnow, L.; and Gonen, R. 2009. Gaming prediction markets: Equilibrium strategies with a market maker. *Algorithmica* 1–40.

Cowgill, B.; Wolfers, J.; and Zitzewitz, E. 2010. Using prediction markets to track information flows: Evidence from Google. Working paper.

Das, S., and Magdon-Ismail, M. 2009. Adapting to a market shock: Optimal sequential market-making. In Koller, D.; Schuurmans, D.; Bengio, Y.; and Bottou, L., eds., *Advances in Neural Information Processing Systems 21*. 361–368.

Das, S. 2005. A learning market-maker in the Glosten-Milgrom model. *Quantitative Finance* 5(2):169–180.

Das, S. 2008. The effects of market-making on price dynamics. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS '08, 887–894. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Gneezy, U., and Rustichini, A. 2000. Pay enough or don't pay at all. *The Quarterly Journal of Economics* 115(3):791–810.

Hanson, R., and Oprea, R. 2009. A manipulator can aid prediction market accuracy. *Economica* 76(302):304–314.

Hanson, R. 2007. Logarithmic market scoring rules for modular combinatorial information aggregation. *Journal of Prediction Markets* 1(1):3–15.

Hanson, R. 2008. Insider trading and prediction markets. *Journal of Law, Economics, and Policy* 4:449–463.

Loewenstein, G. 1999. Experimental economics from the vantage-point of behavioural economics. *The Economic Journal* 109(453):25–34.

Luckner, S., and Weinhardt, C. 2007. How to pay traders in information markets? Results from a field experiment. *Journal of Prediction Markets* 1(2):147–156.

Othman, A., and Sandholm, T. 2010. Automated market-making in the large: the Gates-Hillman prediction market. In *Proceedings of the 11th ACM Conference on Electronic Commerce*, 367–376.

Othman, A.; Sandholm, T.; Pennock, D.; and Reeves, D. 2010. A practical liquidity-sensitive automated market maker. In *Proceedings of the 11th ACM Conference on Electronic Commerce*, 377–386.

Pennock, D., and Sami, R. 2007. Computational aspects of prediction markets. In Nisan, N.; Roughgarden, T.; Tardos, E.; and Vazirani, V. V., eds., *Algorithmic Game Theory*. Cambridge University Press.

Pennock, D. 2004. A dynamic pari-mutuel market for hedging, wagering, and information aggregation. In *Proceedings of the 5th ACM Conference on Electronic Commerce*, 170–179.

Servan-Schreiber, E.; Wolfers, J.; Pennock, D.; and Galebach, B. 2004. Prediction markets: Does money matter? *Electronic Markets* 14(3):243–251.

Wolfers, J., and Zitzewitz, E. 2004a. Five open questions about prediction markets. Stanford GSB Working Paper.

Wolfers, J., and Zitzewitz, E. 2004b. Prediction markets. *Journal of Economic Perspectives* 18(2):107–126.