DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

INFORMATION AGRREGATION MECHANISMS: CONCEPT, DESIGN AND IMPLEMENTATION FOR A SALES FORECASTING PROBLEM

Charles R. Plott

Kay-Yut Chen
Hewlett Packard Laboratories

SOCIAL SCIENCE WORKING PAPER 1131

March 2002

Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem

Kay-Yut Chen
Hewlett Packard Laboratories
and
Charles R. Plott
California Institute of Technology

ABSTRACT

Information Aggregation Mechanisms are economics mechanisms designed explicitly for the purpose of collecting and aggregating information. The modern theory of rational expectations, together with the techniques and results of experimental economics, suggest that a set of properly designed markets can be a good information aggregation mechanism. The paper reports on the deployment of such an Information Aggregation Mechanism inside Hewlett-Packard Corporation for the purpose of making sales forecasts. Results show that IAMs performed better than traditional methods employed inside Hewlett-Packard. The structure of the mechanism, the methodology and the results are reported.

Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem

Kay-Yut Chen
Hewlett Packard Laboratories
and
Charles R. Plott
California Institute of Technology

1. INTRODUCTION[1]

Information and knowledge in social systems frequently exist only as dispersed insights, local data, opinions and guesses and their importance to the operation of social systems, regardless of the subjective or "soft" form, is widely recognized. However, exactly how such information and knowledge finds its way to useful social purposes and whether or not the information that exists in the system is fully used, appears to be highly dependent on the nature of organizations and institutions that support its transfer and use. The variety of processes found in organizations, such as committees, polling processes, networks of contacts, reporting, etc. developed for the purpose of improving information flow suggests both the importance and complexity of the task.

The work reported here finds its early motivation in the classical work of Hayek (1948), who suggests that prices in naturally occurring, free markets make important contributions to information transmission in economies. More precise suggestions of exactly the type of markets that

[1] The financial help of the National Science Foundation and DARPA are gratefully acknowledged. Effort sponsored by the Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory, Air Force Material Command, USAF, under agreement number F30602-00-2-0623. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Defense Advanced Research Projects Agency (DARPA), the Air Force Research Laboratory, or the U.S. Government.

might be successful and how it might take place are found in the theoretical literature of rational expectations. The operational development of theory, the exploration of the possibilities that real markets present and the testing of various market architectures that might support information aggregation are strictly the product of experimental economics. This paper takes the methodologies of experimental economics one step further and reports on the implementation and use of market processes that were designed explicitly for the purpose of performing the task of accumulating and organizing information that is held in a widely dispersed and subjective form.

As is reflected in the title of the paper, the analysis is devoted to the design and implementation of an "Information Aggregation Mechanism". The freedom of design removes any need to restrict the inquiry to a particular institution or some class of institutions that have evolved naturally. The question raised by the research is whether or not the capacity of any competitive processes can be harnessed and developed into an information aggregation tool that can be of use to manage business. Indeed the inquiry can be extended to any type of system that might successfully perform the task. The particular Information Aggregation Mechanism developed and implemented here is closely related to mechanisms suggested by the theoretical and experimental literature on markets.

Economists have long understood that, in theory, the prices in properly designed markets reflect the collection of all the information possessed by all the traders about future events. Almost daily the business pages interpret market behavior as having anticipated events or as having integrated a complex pattern of information into the price. Intuitions that are supportive of such ideas are readily available. Reflecting on every day notions about the way that people learn from observing other people can form a common sense impression of how an aggregation mechanism might work. For example, the observation of many people eating at a restaurant suggests that it is a good place to eat and increases the likelihood that someone will try it. Or, if a crowd is observed looking at something then there is a propensity for additional people to look, thinking that the crowd might be seeing something of interest. The actions of the crowd suggest they know something and others instinctively incorporate this possibility in their own information base. At a more formal level the idea is seen in stock market reports that the markets "anticipate" events like inflation, individual company reported profits, reports on the economy, weather reports, prospects of disruptive events like strikes or wars, etc. The "insiders", those with bits and pieces of information, those with good "intuition" about events are registering their beliefs through their actions in the markets. That is, according to interpretation the markets are like a vacuum sweeper, collecting and aggregating information that is otherwise highly decentralized and privately held. The transformation of these common sense properties into the principles of a science and the design of specific types of processes to enhance such features have resulted in what we will call Information Aggregation Mechanisms (IAM).

A joint research project between Caltech and Hewlett-Packard Laboratories was initiated in 1996 to investigate the possibilities of implementing an Information Aggregation Mechanism. Many business examples share the following characteristic: small bits and pieces of relevant information exists in the opinions and intuition of individuals who are close to an activity. Some examples are supply chain management issues, demand forecasting, new product introduction, and supply uncertainties. In many instances, no systematic methods of collecting such information exist. In these cases very little is known by any single individual but the aggregation of the bits and pieces of information might be considerable. For instance, it is extremely difficult to combine subjective information such as the knowledge of a competitor's move with objective information such as historical data. In a perfect world, with unlimited time and resources, a user of such information could personally interview everyone that might have a relevant insight but such luxury does not exist. Gathering the bits and pieces by traditional means, such as business meetings, is highly inefficient because of a host of practical problems related to location, incentives, the insignificant amounts of information in any one place, and even the absence of a methodology for gathering it. Furthermore, business practices such a quotas and budget settings create incentives for individuals not to reveal their information. The principles of economics together with new technologies that exist for creating markets and related mechanisms suggest that in might be possible to develop a new approach that avoids many of the practical problems.

The research posed here asks many fundamental questions that can only be answered by field demonstration. Abstract theory or laboratory experiments might help but in the end cannot provide answers. What types of mechanism might work? How can it be deployed within a business or other types of organizations? Does it really work? Of course there are many scientific issues that one would like to resolve before an application is attempted. How is the performance of the system related to the psychology and decision biases of individuals? How can one deal with incentive problems in which individuals might large incentives to conceal or misrepresent what they know? What rules and mechanisms might be needed for different underlying information structures? If markets are thin or the number of participants few, how will the performance of the system be affected? How can we find the people with the relevant information and how do we know that they knew something of relevance anyway? The mechanisms are supposed to aggregate information that is there and not create it from nothing. If the participants know nothing, the mechanism will produce nothing.

Some questions have been answered with laboratory experimentation. The proof that properly designed markets can aggregate information has existed in the experimental economics literature for over a decade.[2] Experimentalists have discovered that markets populated by humans can indeed perform in a manner suggested by the theory but that ability is closely related to the market organization.

The feasibility of deploying an IAM in a business environment is suggested by the Iowa Electronic Markets, and comparisons of major features of the Hewlett Packard IAM with the Iowa Electronic Markets (IEM) are of interest. While there are important overlaps, there are substantial differences in the environments for which these mechanisms were applied. There are also differences in the methodologies themselves. The IEM are focused on events that are observable to a broad based population at large, such as election outcomes, certain stock prices, or the

[2] The experimental demonstration is first found in Plott and Sunder (1982, 1988). This early paper demonstrated that the ability of markets to aggregate information is sensitive to the market architecture. In particular, this early work demonstrated that compound securities are not as reliable as indicators as a complete set of state dependent instruments. The conditions under which a single compound security is reliable are isolated in Forsythe and Lundholm (1990) The need for selecting proper instruments is underlined by demonstrations of markets that can equilibrate at patterns that are not fully revealing of information such as cascades (Anderson and Holt, 1997; Hung and Plott, 2001) or misleading such as mirages (Camerer and Wiegelt, 1991) or bubbles (Smith et al, 1988; King et al. 1993; Porter and Smith, 994; Lei et al, 2001). In fact, some types of market organization facilitate no information aggregation at all as is the case of the winners curse in sealed bid auction markets (Kagel and Levin, 1986; Lind and Plott, 1991). See Sunder (1995) for a summary, or aspects of search (Sunder, 1992).

actions of the Federal Reserve. For example, in the Iowa Electronic Markets, participants are allowed to buy and sell “shares” of candidates in U.S. and even foreign elections. In such cases, it is not clear that the information to be captured involves any substantial degree of aggregation in the sense that the concept is applied for the operation of the IAM. In the case of public events, much of the relevant information may already be public knowledge. It is not clear that IEM participants have specific, specialized information that is not available for the general public. Many public reports, polls and summaries exist in the press and obviously affect IEM activities. Participants in the IEM have self-selected for participation, which has no specific bearing on whether they have private information or not. Thus, the predictions of the IEM could reflect an effective and sophisticated system of polling, a collection of personal intentions, a coordination mechanism of public information, or a combination of polls. While related, such patterns of underlying information can be differentiated from a pattern of private information (beyond the personal intention to vote) that resides in small amounts across the population.

By contrast, only a relatively small number of people were chosen for participation in the Hewlett Packard IAM. They were selected specifically from different parts of the business operation because they were thought to have different patterns of information about the targeted event. These patterns of information, including market intelligence, specific information about big clients, and pricing strategies, were in need of aggregation. In addition, there were no public summaries of information available to the participants during the operation of the IAM. The official forecasts were not known until after the IAM closed. In fact, antidotal evidence suggests that the activities on the IAM were used as inputs to HP official forecasts in more than one occasion.

The Iowa Electronic Markets were thick, with many participants operating over a long period of time while the markets in the HP IAM were thin and operated over very short periods. Such differences lead to procedural and market architecture differences. Differences in the nature of the prediction task also lead to different market instruments. The primary predictions of the IEM are point predictions, such as vote shares in political events, which are designed to compare well with polling techniques. The overlap between the IEM and the HP IAM is greatest in the IEM “winner take all” markets. Such markets are state contingent securities and are based on the same principles as the HP IAMs reported here. This overlap between the two substantially different types of exercises can further attest to the basic principles' robustness in widely varying environments.

2. THE TASKS

A total of twelve predictions were performed over a period of three years. These are listed in Table 1. Included in the table are (i) the events to be predicted, (ii) when the IAM was conducted ,(iv) the number of participants, (v) the duration over which the information aggregation mechanism operated and (vi) some information about the structure of the mechanism.

The first event prediction attempted was the level of profit sharing bonus in a particular half year. The bonus is calculated based on a pre-defined formula of HP's profit. A total of eight possibilities were defined, e.g. below 2%, between 2% and 3%, etc. and a state contingent market was opened for each.

All other events predicted were related to sales of products. Predictions for eight products were conducted, which are labeled {A,B,C,D,E,F,G,H} for purposes of discussion and analysis. As can be seen in the table predictions were performed more than once for some of these products and only once on others. In some cases dollar sales were predicted for some month and in other cases it was the number of units sold.

Typically, the prediction was for monthly sales for a month three months in the future, with the exception being the first two exercises for which the predicted event was one month ahead. In all cases the information was gathered for a week with the markets being open during lunch and in the evening every day. Management did not want participants being preoccupied with the task during the working day when pressing issues needed attention.

3. THE MECHANISM: Technical Issues

Technical issues are of two types. The first issue is related to the instruments that will be used in the markets. Exactly what will be the items that are bought and sold in the markets? The second issue is related to the market mechanism, the technology for making bids, asks and trades.

The primary choice of instruments was between a single compound security, which paid a dividend in proportion to the level of sales, if sales are the item to be forecast, and multiple, state contingent contracts. Because it is known from experiments that single compound securities can have difficulty with information aggregation Plott and Sunder (1988) a decision was made to use a complete set of state contingent contracts.

Arrow-Debreu securities were chosen as the instruments because of the aggregation success that they have exhibited in the laboratory (See Plott, 2000) for an example of the types of laboratory experiments that had been used to explore this issue.) The space of possible outcomes was partitioned into a finite number of subsets. Each subset was "tied" to a security. After the final outcome was revealed, the security, which contained the final outcome, was determined. This "winning" security paid off a fixed amount. All other securities paid nothing.

Since all the events chosen to be predicted lie on the positive real line, securities were constructed by partitioning the real line into about 10 or so (exact number depends on the event) intervals. Each interval was given a name and with each interval there was an associated security with the same name that traded in a market with that name. Thus the interval 0-100 would be associated with a security named 0-100 that traded in a market named 0-100. The interval 101-200 would be associated with a security named 101-200, etc. If the final outcome fell in an interval, the corresponding security would pay, say, one dollar per share at the end of the experiment. All other securities would pay nothing.

Of course the exact amount of the payoff per share was a parameter that could be changed but in the business exercises reported here it was always $1.00 per share. A higher payoff per share would place more value on the share but the payoff per share interacts with the total cost of the exercise and the potential volume of trades and related market liquidity.

Each participant was given a portfolio of shares in markets and cash. In some exercises all participants were endowed equal shares in all securities. In other exercises participants were endowed with shares in every other security, alternating which security was first across participants. The unequal distribution of endowments was used to encourage trading by attempting to make sure that the initial endowments of securities did not approximate the ultimate equilibrium. Of course the total number of shares and cash distributed determined the overall cost of the exercise.

The market mechanism employed to support the markets was the web based, double auction markets of the Marketscape software, which was developed at the Laboratory of Economics and Political Science at Caltech. All the markets for an event were organized on a single web page for easy access. As can be seen from the screen shot in Figure 1, each of the different markets occupies a line on which the bid, ask and last trade price are all displayed. Links to a complete time series of trades in their graphical or as raw data were available. Links also existed to HP data bases, which allowed participants to review data held by HP. A participant could enter a buy offer, a sell offer or acceptance of an offer through the web form on the page. Orders were compared to the other side immediately. If a trade was possible, it was executed and if not the order was placed in an order book. The best offers were listed on the main market web page. The whole book of offers was available for each market at the click of a button.

Participants were located in a diverse geographical area. Some may have been traveling during the experiments. Participation was anonymous. However, each participant was assigned a subject ID number for each experiment. During the experiment, the subject ID number of the person who made offers and transactions were public knowledge. Participants had the ability to track behavior of other subjects within the same experiment if they wished to.

4. THE MECHANISM: Business Issues

Implementation of a market institution in a business environment offers different challenges than implementation of the same institution in a laboratory environment. The major issues are participant selection and motivation together with other business constraints.

Selection of participants is of primary importance. In an academic experiment, participants are given appropriate amount of information by virtue of experimental controls and procedures. Therefore, there is no issue of selecting participants for what they might know since all of them, by design, will have appropriate information. The only issue is whether or not a mechanism is aggregating information that is known to exist.

By contrast, in a business environment, participants need to be selected carefully. On one hand, it is not desirable to "miss" a person with much information. On the other hand, it might not be efficient to include many people without any relevant information. Indeed, little is known theoretically about the information size relative to the market that might be required for effective information aggregation to take place.

A second issue is ensuring participation through proper incentives and timing. Even if the right people are identified to participate, it is not a foregone conclusion that the appropriate amount of participation would be forthcoming. The opportunity costs of doing something other than participating in the forecasting exercise can be very high for the business people. Consequently, a level of incentives much higher than that in an academic experiment is often required. Secondly, scheduling market sessions is also problematic. On one hand, it is desirable to have a schedule (for example, 24 hours for a week) to minimize conflict with other activities. On the other hand, it is not desirable to leave market open for long periods because participant will often find a lack of activities in the markets and thus lose interests. Third, it is known from laboratory experiments that the ability of markets to predict increases as the participants have experience with the markets and with each other. Thus, the incentives and the schedules should be such that participants enjoyed participation and were willing to repeat the experience with several prediction tasks.

Business constraints operate in different dimensions. First, there may be some hesitation to engage employees in an exercise in which they might lose money. Certainly this was the case in HP. Thus, we had to provide a small amount of cash to each participant before the market sessions. This, together with the budget determined together with the HP management, constrained the amount of stakes a participant can have in the market and affects incentives to trade. Secondly, the business implementation had to offer possibly useful information. In particular the horizon of the predictions was important. Typically, forecasts are not valuable if they are made with horizons less than 3 months. Therefore, market sessions need to be conducted 3 months before the event to be predicted.

5. PROCEDURES

The experiments were conducted with three different HP divisions. Typically, around 20-30 people signed up for the experiments. Business participation was limited to marketing and finance organizations. An addition of around five subjects was recruited from HP Labs in each experiment. The reason of adding HP Labs subjects, who had little or no information about the predicted event, was to increase liquidity in the markets. Laboratory experiments have suggested that a small number of uninformed participants provide both market liquidity and a function of adding “consistency” to the market through a process of “reading” and “interpreting” the actions of others.

The markets were typically open for about a week for each event. Within the week, there were additional restrictions on when the markets were open, as was discussed elsewhere. Trading was done through at a web server located at Caltech. The subjects were geographically dispersed in California.

Some effort was made to make participation anonymous. However, since most of our subjects worked for the same organization, we would expect normal interactions amongst the subject pool during the experiments.

Each subject was given a 15-20 minute one-to-one instruction session from one of the experimenter. The experimenter explained the structure of incentives and the market mechanism as well as the web interface in details. Contact information was also given to each subject. They were encouraged to call if they encountered difficulties. In addition, the participants were told the goals of the experiment and were told that their participation was important for HP business. While this procedure is atypical of laboratory experiments the goal of the exercise presented here was to design and implement a mechanism that works and the instruction about the purpose was thought might help.

At the end of each exercise, the subjects were paid by checks or by the internal HP reimbursement procedures.

6. FORESCASTS

The distributions of prices computed as the last 50% of trades in a market are shown for all markets for all events forecast in Figures 2A thru 2L. The actual outcomes, the IAM predictions as well as the official forecasts (if available) are also indicated on the figures. For example, event 2 (figure 2B) shows the IAM distribution to have a single peak around 230. The actual outcome (the dotted line) is at 220 while the official forecast is at 249. The IAM prediction (solid vertical line) comes in at 230 in the middle between the actual outcome and the official forecast.

Visual inspection shows that the IAM predictions are closer to the actual outcomes than the official forecasts in events 2,3,4,5,6 and 9. It is also worth noting that in event 3,4 and 5, the IAM predictions are so close to the official forecasts that they are practically on top of one another.

The forecasts (predicted sales) resulting from the markets are contained in Table 2. The table also contains the actual sales that resulted in the period for which a prediction was made. For example the outcome event 2 was a sales level of 220 as compared to the official HP forecast of 249, which had a 13.182% error. The IAM prediction for event 2 using the last 50% of trades was 230.059 with a percent error or 4.572.

A discussion of the predictions derived from the IAM is in order. The key issue in interpreting the results is to find a reasonable way to derive predictions from the market data. Theoretically, the prices of an efficient market should sum to the payoff of the Arrow-Debreu securities and be proportional to the probabilities of the states conditioned on the information in the market. However, in our experiments prices did not always sum to the right amount or stay on a stable level. Thus, the price used as an IAM forecast used to calculate states probabilities becomes an issue.

Several options present themselves as predictive statistics. We chose the volume-averaged transaction price of the IAM markets as the measure. The averages were taken over a subset of the trades towards the end of the experiments. The rationale being the market achieved some sort of equilibrium towards the end as suggested by laboratory experiments. Since no objective criterion exists for choosing the number of trades to include, different percentages were used and the results seem to be robust with respect to that.

The model employed assumes that for each interval the event was equally likely within the interval. That assumption when joined with the probabilities of each interval calculated from the IAM market prices, enables a calculation of an expected outcome. Since all events reside on the real line with no upper bound, all but the upper most intervals were bounded and thus posed no issue in this calculation. The upper most interval, however, so to deal with this asymmetry two different approaches were tried. The first approach ignored the last interval completely since for most experiments, the probability placed on it was small. The second approach, assumed all the probability in the last interval was placed on its lower bound. Results were robust with respect to how expected value (of sales) was calculated.

The second major issue is the choice of an appropriate benchmark to measure the performance of the IAM predictions. In a business environment information held by all individuals is not known. It is impossible to benchmark the IAM predictions with respect to the total information available as done under laboratory conditions. The only benchmark available is the HP official forecast. The HP official forecast is a logical benchmark since it represents the "belief" of HP but it does have limitations. For example, the HP official forecast is not accompanied by assumptions about the distribution of accuracy. In fact, the official forecast is not only a forecast, it is a management tool through which quotas are measured and compensation paid. This dual role of the forecast means that it is expected and management and divisions coordinate on the assumption that the forecast will be met. Furthermore, HP official forecasts were only available in 8 out of 12 experiments that we have conducted.

7. RESULTS

The overall pattern of results indicates that the IAM prediction is a considerable improvement over the HP official forecast. The IAM predictions relative to official HP forecasts are found in Table 2. The actual outcome, HP official forecasts and their absolute percentage errors are listed. IAM predictions calculated with three different methods are also reported with their absolute percentage errors.

Table 3 reports the comparison statistics between HP official forecasts and IAM predictions. The absolute percentage errors of the HP official forecasts, as well as those of IAM predictions calculated with six different methods are listed. T-tests were used to determine whether the HP official forecasts have higher errors than each of the six IAM methods. The p-values of these tests are reported in the same table.

These form the basis of the first result.

RESULT 1: Market predictions based on IAM prices outperformed official HP forecasts.

Support:

In 6 out of 8 events for which official forecasts were available the IAM predictions are closer to the actual outcome than the official forecast. T-tests also show that the absolute % errors of the official forecasts are higher than that of the IAM predictions. The hypothesis that the error of the official forecast is less than or equal to the error of the IAM prediction can be rejected. The result that the IAM forecasts are more accurate is robust with respective to different calculation methods. Table 3 summarizes all forecasts and predictions errors and the t-statistic for each of the following definitions of volume averaged transaction prices that are used as the probability (after normalized) of a particular interval. The IAM prediction is the interval in which sales are predicted to fall. Thus, the predicted interval is sensitive to the probabilities attached to intervals and the interpretation attached to the highest interval. Robustness was checked for the following assumptions.

a. using last 40% of the trades to calculate the volume averaged price
b. using last 50% of the trades to calculate the volume averaged price
c. using last 60% of the trades to calculate the volume averaged price
d. ignoring the last interval to calculate an expected value of the predicted variable
e. assuming all the probability of the last interval concentrates on its lower bound to calculate the expected value of the predicted variable

As Table 3 reveals, regardless of the definition used the results remain essentially unchanged.•

The first result is focused only on point predictions. However, the IAM also produces probabilistic information about the likelihood of possible outcomes. Prices in a market can be interpreted as the probability that sales will fall in the interval represented by the market. Instead of producing a point prediction the IAM produces a whole probability distribution. The next result deals with the question of whether the probability distributions derived from prices are consistent with actual outcomes.

The intuition for the test is as follows. Each event was potentially generated by a different probabilistic process. However, if the IAM process recovers the true distribution of the events, the derived distribution, unique to each event, should be consistent with the actual outcome. Thus, ten percent of the outcomes should fall below the 10% quantile of the derived distribution. Twenty percent of outcomes should fall below the 20% quantile of the derived distribution and so on.

The mathematical foundation for the test uses the following argument and resulting proposition. Let $x \in R$ for some finite set R of the real numbers and let H(x) be the probability that a random variable *x* takes values less than or equal to x under the model H. Let F(x) be the true probability that the random variable *x* takes a value less than or equal to x. Define a function $y = H(x)$ and consider the inverse correspondence $H^{-1}(y)$ and a related function $\mathit{H}^{-1}(y) \in H^{-1}(y)$ which takes a single value for each y^* such that $H^{-1}(y^*)$ is not a single element. Such a selection can always be done with finite sets and with larger sets the Axiom of Choice can be applied. Of course $y \in [0,1]$. Let P(y) be the probability of y, so by the assumption that F(x) is the true probability and by using the definition of $\mathit{H}^{-1}(y)$ we have $P(y) = F(\mathit{H}^{-1}(y))$.

Furthermore, if F(x) = H(x) we can define a function $\mathit{F}^{-1}(y)$ such that $\mathit{F}^{-1}(y) = \mathit{H}^{-1}(y)$. Now if $H(x) = F(x)$ then substituting we have the following proposition for $y \in H(x)$ for some x.

PROPOSITION: $P(y) = F(\mathit{F}^{-1}(y)) \equiv y$. That is, P(y) is the uniform distribution between 0 and 1.

Now assume there are N random variables labeled x_1 through x_n. Let $H_i(x_i)$ be the probability distribution of x_i under the IAM prediction. Let $y_i = H_i(z_i)$ where z_i is the observed actual outcome of x_i.

A direct result of the proposition above is that for all i, y_i is distributed uniformly between 0 and 1. Furthermore, since these events are independent, y_i is also i.i.d.

Thus, testing whether y_j is distributed uniformly is an indirect test of whether the predicted distributions (H_j) are consistent with the true distributions (F_j).

RESULT 2: The probability distributions calculated from market prices are consistent with actual outcomes.

Support:

A Kolmogorov-Smirnov goodness of fit test was conducted on the y_j with the null hypothesis that the y_j is distributed uniformly in the interval [0,1]. The test was conducted with 12 data points (12 events). The ks-statistics is 0.1743 with a p-value of 0.80. The null cannot be rejected at any level of reasonable significance.

Thus, one can reasonably conclude that the distributions derived from the IAM process are consistent with the underlying distributions that generated the events. •

The next result is another test of the accuracy of the IAM relative to official forecasts. The question posed is the nature of any improvement over official forecasts. Does the IAM suggest adjustments in the official HP forecast that are in the "right direction"? The next result reports that the answer is "yes".

RESULT 3: The IAM makes accurate qualitative predictions about the direction that the actual outcome will occur (above or below) relative to the official forecast.

Support:

We use the following procedure to predict whether the actual outcomes will be above or below the official forecasts. First we calculate the predicted distribution based on last 50% of the trades in each market. Then we look at the point of the official forecast and determine whether the distribution is "skewed" (have more mass) to the left or the right. If the distribution has more mass on the right, then we predict that the actual outcome is going to be higher than the official forecast. Otherwise, the actual outcome is going to be lower than the official forecast.

Using this method, the actual outcomes are consistent with the prediction 8 out of 8 events that had official forecasts. Please see Table 4 for a summary of the predictions. •

In addition to the results stated above several observations are listed below. These are related to phenomena that are either of importance to theory or are of interest for establishing a correspondence with laboratory experimental data.

OBSERVATION 1: Theoretical arbitrage profits existed.

In all the experiments, prices summed to be greater than the winning payoff. In theory, it violated the no arbitrage conditions. However, to take advantage of the arbitrage conditions, individuals needed to execute multiple trades when fluctuations of prices were substantial. Thus, although violations of theoretical arbitrage conditions were observed in the experiments, it is likely that there were actually no practical arbitrage opportunities. However, this alone only explains why the sum of prices can be substantially away from the efficient point (summed up to be the winning payoff). It does not explain why in all 12 experiments the sum of the prices was always above the winning payoff.

OBSERVATION 2: No significant information about the state was identified in the dynamics observed during the course of the experiments.

For each experiment, trading data were divided into 10 subsets, each containing 10% of the trades in sequence. The first subset contained the first 10% of the trades. The second subset contained the second 10% of the trades and so on. Market prices were calculated using trades in each subset and point predictions were calculated for each subset based on the prices. No significant trends in the sequences of predictions are observed.

One speculation is that information aggregation occurred fairly early. Under that hypothesis, any variations in the predictions during the course of the experiments were due to drift around equilibria.

OBSERVATION 3: The number of sell offers exceeded the number of buy offers.

In all experiments, there were more sell offers than buy offers. This phenomenon is typical when prices are going down. This is intuitively consistent with what we know about information markets. In the early stage of the markets when little information was reflected, prices were around the same levels in all possible outcomes. As the market progressed and aggregated more information, prices associated with most outcomes (except the few likely ones) decrease. We observed more sell offers than buy offers because prices associated with most of the outcomes were decreasing.

7. CONCLUDING REMARKS AND BUSINESS QUESTIONS AND ISSUES

The goal of this research is to take known information aggregation mechanisms, attempt to deploy them within a real business environment and determine whether or not they will "work". The IAM chosen was a set of markets consisting of a complete set of Arrow-Debreu securities. Laboratory experiments suggested that a mechanism of this form would be successful. The IAM was successfully implemented in several Hewlett-Packard business divisions. Numerous roadblocks were encountered, such as scheduling conflicts and a lower than ideal level of participation, which are non-issues in academic experiments but are extremely important in the business environment.

The results are encouraging. Not only did the IAM market predictions consistently beat the official HP forecasts; the outcomes predicted are consistent with the probabilistic predictions of the IAM.

The methodology has generated interest in the HP business community, despite a number of research issues, because of advantages this method has that other forecasting methods do not have.

(i) The IAM is flexible. It can be used to aggregate any type of information possessed by different people. It involves a natural methodology for quantifying subjective, qualitative, information and giving weights to the opinion of different people for the purpose of information aggregation. The task is performed giving not only a point forecast but also a complete probability over the range for which the value of some unknown variable is to be predicted.

(ii) The methodology is scalable by number of participants, timing of participants and location of participants. There are no practical limits to the number of people that can participate. With markets conducted over the Internet, hundreds and even thousands of people can participate either at the same time or at different times. Traditionally, businesses collect and aggregate information through a process of meetings, which not only limits the number of participants but also the time frame for information collection.

(iii) The methodology tends to be incentive compatible. Incentives to hide information, misrepresent information or simply ignore requests for information are either eliminated or limited. Furthermore the markets are designed to give incentives to the participants' to acquire information about future events and use this information wisely in the market.

Obviously this is only the first step in turning special designed economics mechanisms into forecasting devices. Once feasibility is established a number of questions naturally arise. Do alternative mechanisms exist that will aggregate information better? Does the structure of information matter and if so, can an appropriately designed IAM take advantage of that? Can an IAM not only produce a prediction but also simultaneously help management ascertain which participants have information. That is, can it be designed to attract those with good information and discourage those with bad information? Theory and experiments suggest that improvements along such dimensions are possible but the feasibility, when subject to business constraints, remains to be established.

Table 1: Summary of Experiments

	Event to be predicted	Number of active participants	Date [time] of experiment	Experiment Duration	Number of Markets
1	Profit sharing percentage to be announced by upper management	16	10/96 [11:00 AM-1:00 PM; 4:30 PM-8:00AM]	1 week	8
2	Next month sales (in $) of product A	26	11/96 [11:00 AM-1:00 PM; 4:30 PM-8:00AM]	1 week	9
3	Next month sales (in units) of Product B	20	01/97 [11:00 AM-1:00 PM; 4:30 PM-8:00AM]	1 week	9
4	Quarter ahead monthly sales (in units) of product C	21	05/97 [11:00 AM-1:00 PM; 4:30 PM-8:00AM]	1 week	10
5	Quarter ahead monthly sales (in units) of product D	21	05/97 [11:00 AM-1:00 PM; 4:30 PM-8:00AM]	1 week	10
6	Quarter ahead monthly sales (in units) of product B	21	05/97 [11:00 AM-1:00 PM; 4:30 PM-8:00AM]	1 week	10
7	Quarter ahead monthly sales (in units) of product C	24	06/97 [11:00 AM-1:00 PM; 4:30 PM-8:00AM]	1 week	10
8	Quarter ahead monthly sales (in units) of product D	24	06/97 [11:00 AM-1:00 PM; 4:30 PM-8:00AM]	1 week	10
9	Quarter ahead monthly sales (in units) of product E	24	06/97 [11:00 AM-1:00 PM; 4:30 PM-8:00AM]	1 week	10
10	Quarter ahead monthly sales (in units) of product F	12	04/99 [11:00 AM-1:00 PM; 4:30 PM-8:00AM]	1 week	8
11	Quarter ahead monthly sales (in units) of product G	12	04/99 [11:00 AM-1:00 PM; 4:30 PM-8:00AM]	1 week	8
12	Quarter ahead monthly sales (in units) of product H	7	05/99 [11:00 AM-1:00 PM; 4:30 PM-8:00AM]	1 week	8

Table 2: Summary of Forecasts and Errors

Event				IAM Predictions			
				Last Trade	Average Last 60% Trade	Average Last 50% Trade	Average Last 40% Trade
1	Outcome	8.770	IAM Prediction	9.619	9.092	9.259	9.369
	HP Forecast	None	Tail Prob Truncated	0.040	0.038	0.043	0.041
	% error	None	% error	9.683	3.672	5.571	6.829
2	Outcome	220.000	IAM Prediction	234.065	230.136	230.059	230.294
	HP Forecast	249.000	Tail Prob Truncated	0.009	0.008	0.009	0.009
	% error	13.182	% error	6.393	4.607	4.572	4.679
3	Outcome	1152.000	IAM Prediction	1766.399	1814.155	1793.875	1781.017
	HP Forecast	1838.000	Tail Prob Truncated	0.010	0.008	0.008	0.008
	% error	59.549	% error	53.333	57.479	55.718	54.602
4	Outcome	1840.000	IAM Prediction	1612.891	1695.796	1690.102	1683.273
	HP Forecast	1681.000	Tail Prob Truncated	0.008	0.011	0.011	0.011
	% error	-8.641	% error	-12.343	-7.837	-8.147	-8.518
5	Outcome	2210.000	IAM Prediction	1429.839	1526.466	1512.397	1506.579
	HP Forecast	1501.000	Tail Prob Truncated	0.024	0.011	0.011	0.012
	% error	-32.081	% error	-35.301	-30.929	-31.566	-31.829
6	Outcome	128.000	IAM Prediction	91.801	96.985	96.592	95.619
	HP Forecast	90.000	Tail Prob Truncated	0.007	0.010	0.010	0.010
	% error	-29.688	% error	-28.280	-24.231	-24.538	-25.297
7	Outcome	2002.000	IAM Prediction	1828.000	1855.320	1861.382	1867.697
	HP Forecast	2084.000	Tail Prob Truncated	0.008	0.017	0.018	0.019
	% error	4.096	% error	-8.691	-7.327	-7.024	-6.708
8	Outcome	1788.000	IAM Prediction	1728.600	1752.300	1746.033	1755.340
	HP Forecast	1786.000	Tail Prob Truncated	0.008	0.026	0.028	0.021
	% error	-0.112	% error	-3.322	-1.997	-2.347	-1.827
9	Outcome	166.000	IAM Prediction	134.886	126.401	124.748	125.515
	HP Forecast	119.000	Tail Prob Truncated	0.027	0.061	0.073	0.076
	% error	-28.313	% error	-18.743	-23.855	-24.850	-24.389
10	Outcome	30.000	IAM Prediction	15.178	15.017	15.245	15.150
	HP Forecast	None	Tail Prob Truncated	0.148	0.092	0.073	0.072
	% error	None	% error	-49.407	-49.944	-49.184	-49.498
11	Outcome	10.000	IAM Prediction	15.158	15.170	15.308	15.337
	HP Forecast	None	Tail Prob Truncated	0.083	0.082	0.081	0.085
	% error	None	% error	51.583	51.705	53.082	53.368
12	Outcome	17.000	IAM Prediction	15.708	14.991	15.281	15.366
	HP Forecast	None	Tail Prob Truncated	0.085	0.054	0.061	0.064
	% error	None	% error	-7.602	-11.818	-10.112	-9.612

Table 3: T-Test Comparison between Market Forecasts and Official Forecasts

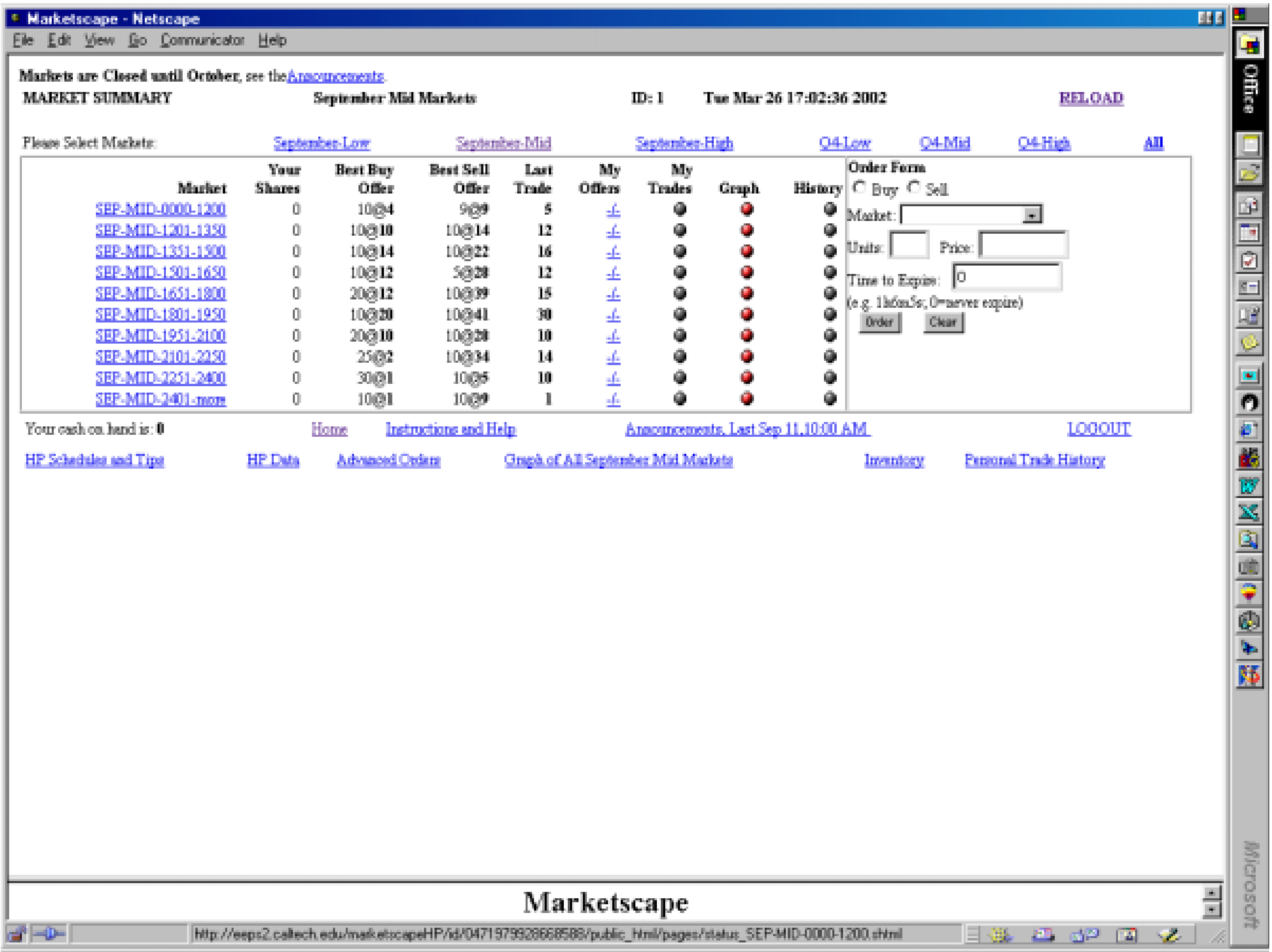
Event	Absolute % errors of HP forecasts	Abs % Errors of IAM Predictions					
		Last Interval Ignored			Last Interval Mass at Lower Bound		
		Average last 60% trade	Average last 50% trade	Average last 40% trade	Average last 60% trade	Average last 50% trade	Average last 40% trade
2	13.18%	4.61%	4.57%	4.68%	5.63%	5.68%	5.80%
3	59.55%	57.48%	55.72%	54.60%	59.25%	57.46%	56.32%
4	8.64%	7.84%	8.15%	8.52%	6.45%	6.77%	7.13%
5	32.08%	30.93%	31.57%	31.83%	29.74%	30.33%	30.48%
6	29.69%	24.23%	24.54%	25.30%	22.94%	23.22%	23.93%
7	4.10%	7.33%	7.02%	6.71%	5.35%	4.91%	4.55%
8	0.11%	2.00%	2.35%	1.83%	1.53%	1.39%	1.00%
9	28.31%	23.85%	24.85%	24.39%	17.55%	17.32%	16.54%
T-test P-value		0.079	0.084	0.071	0.034	0.026	0.022

Random variable x=official error – market error
H0: mean of x=0
Alternate: mean of x>0

Table 4: Predicting "Above" or "Below" Official Forecasts

Event	Cumulative Probability at Official Forecast	Prediction	Official Forecast	Outcome
1	None	N/A	None	N/A
2	86.50%	down	249	220
3	53.79%	down	1838	1152
4	35.62%	up	1681	1840
5	37.46%	Up	1501	2210
6	40.70%	Up	90	128
7	76.33%	down	2084	2002
8	42.92%	Up	1786	1788
9	26.49%	Up	119	166
10	None	N/A	None	N/A
11	None	N/A	None	N/A
12	None	N/A	None	N/A

Figure 1: Home Page for Bidder Participating in the Electronic Market Supporting the Information Aggregation Mechanism



Figures 2A-1L: Distribution Calculated using Last 50% Trades

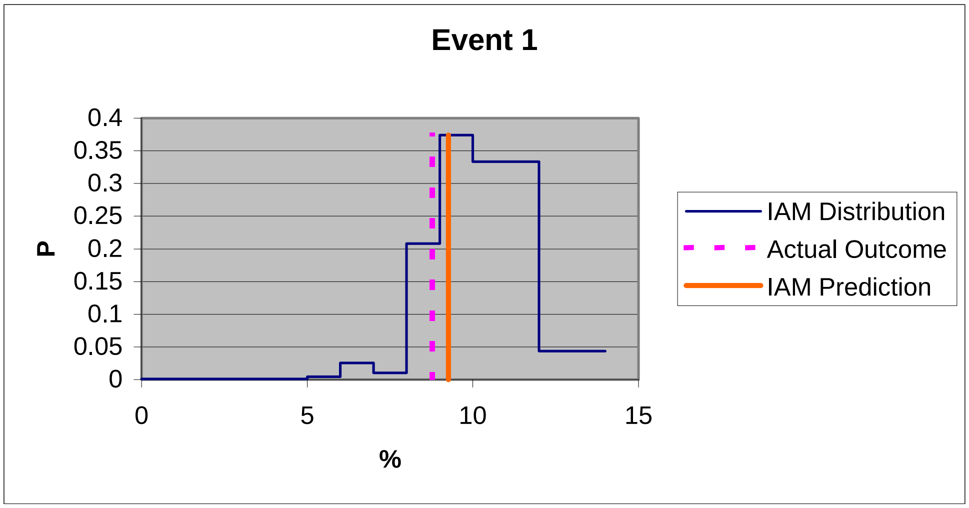


Figure 2A

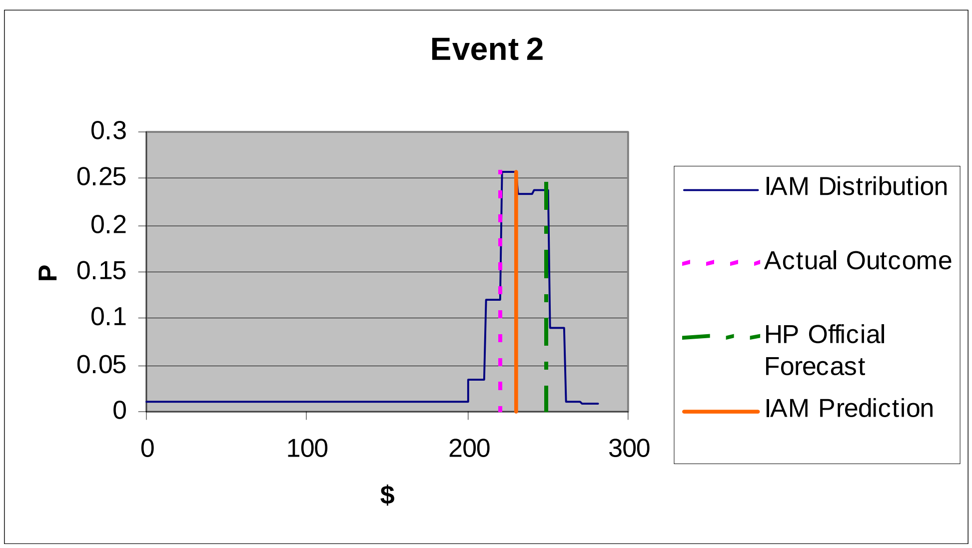


Figure 2B

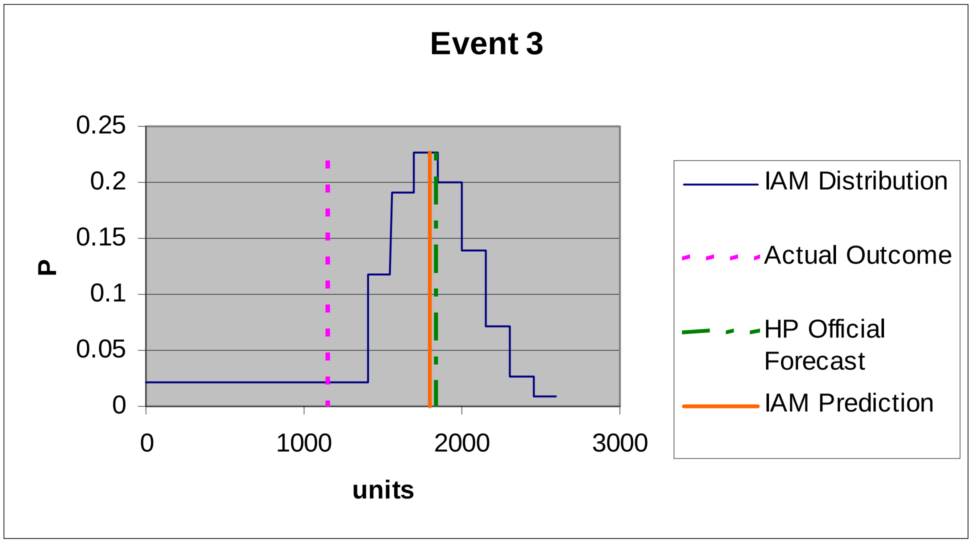


Figure 2C

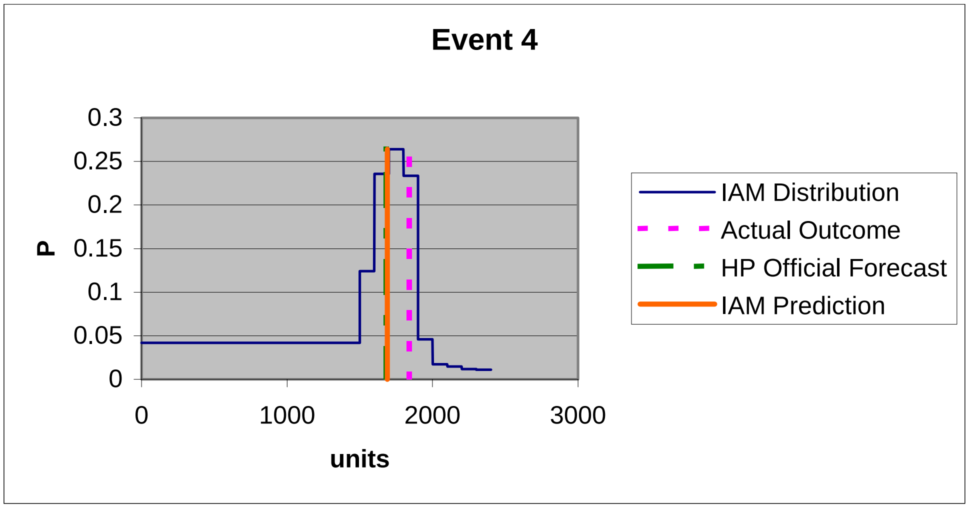


Figure 2D

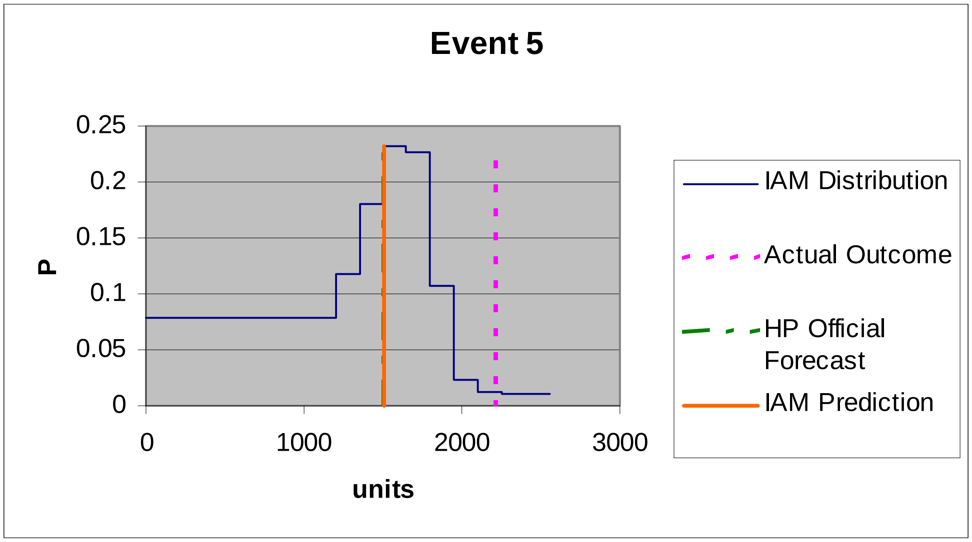


Figure 2E

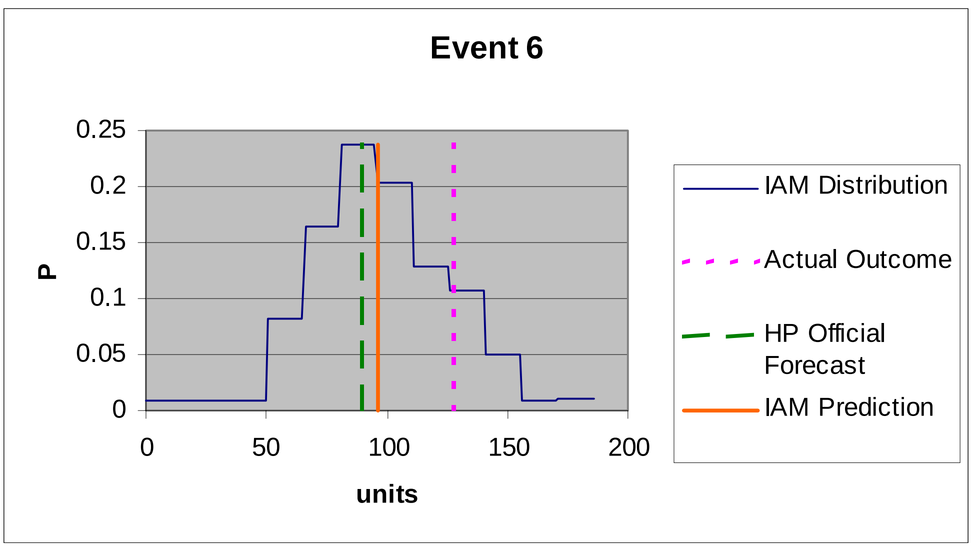


Figure 2F

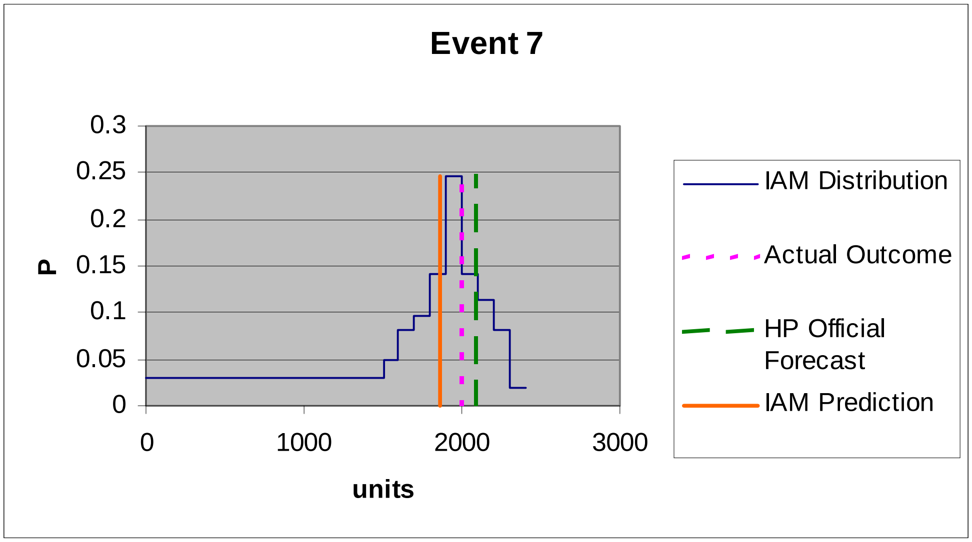


Figure 2G

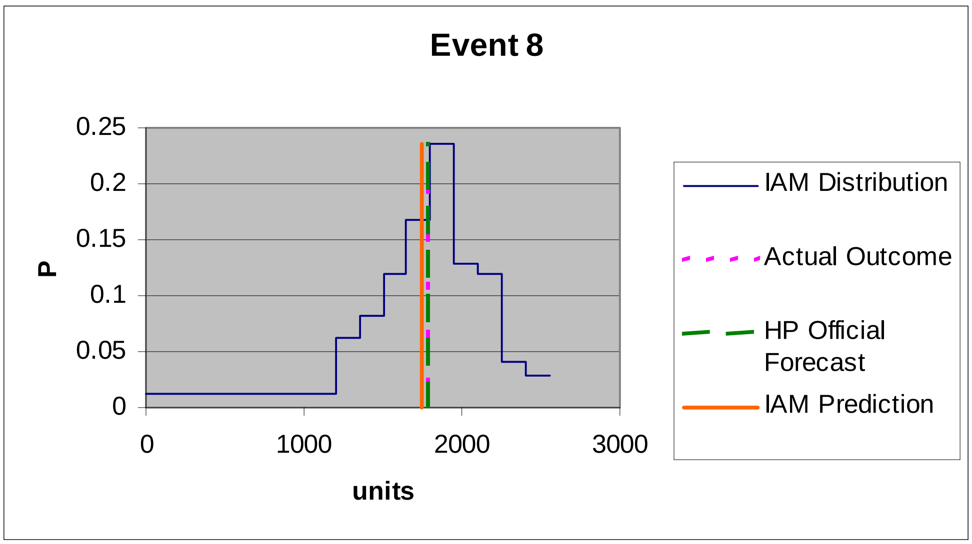


Figure 2H

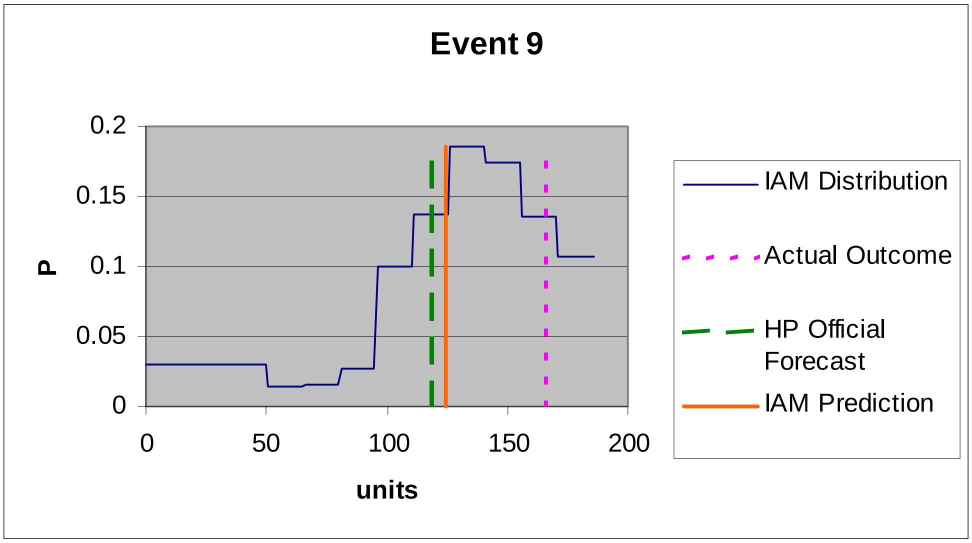


Figure 2I

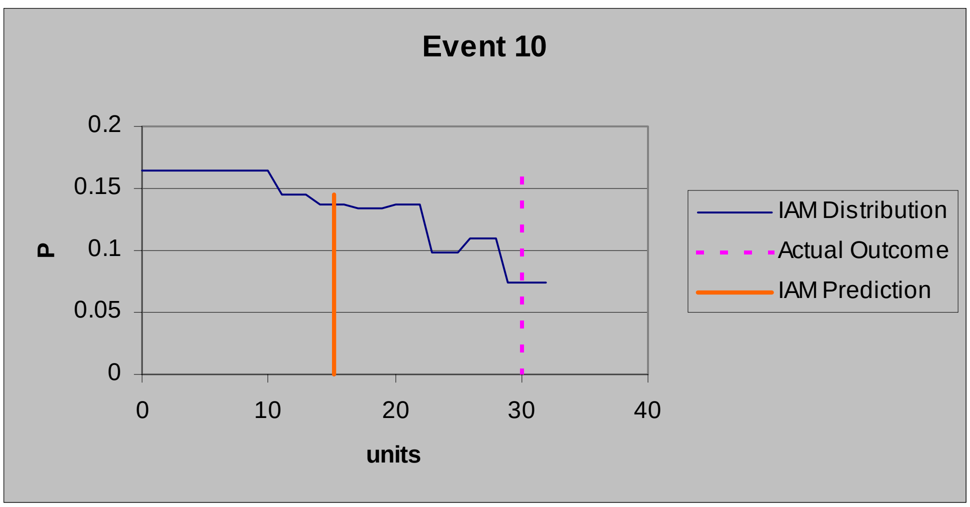


Figure 2J

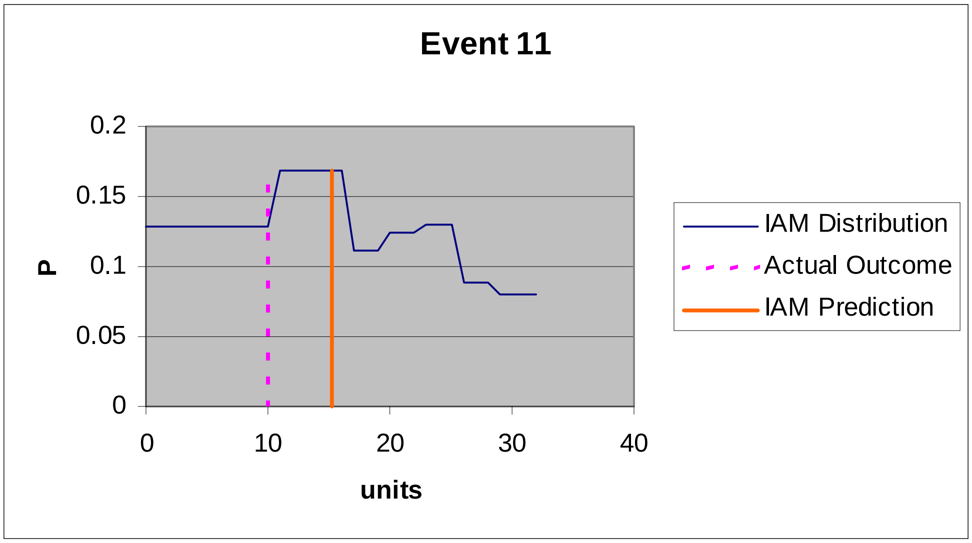


Figure 2K

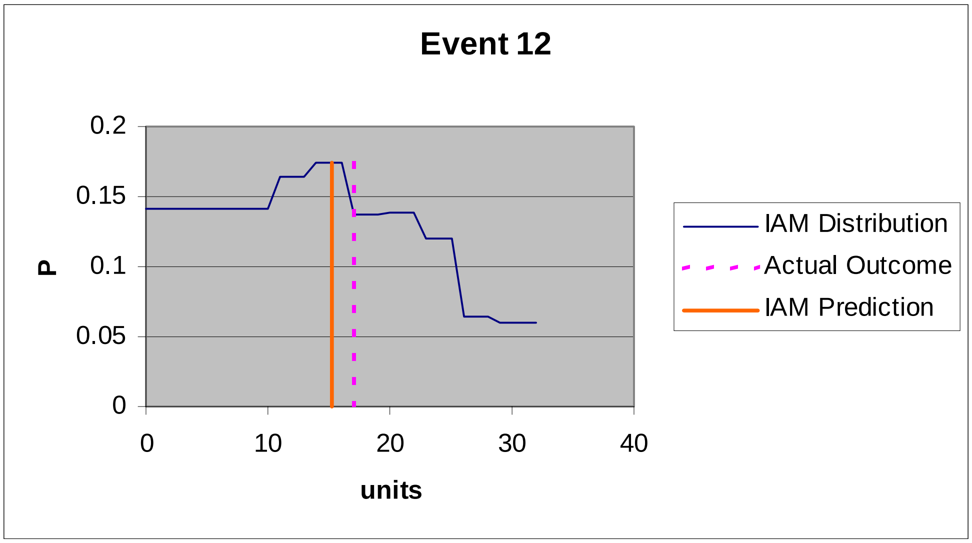


Figure 2L

REFERENCES

Anderson, Lisa R. and Charles A. Holt, "Information Cascades in the Laboratory." *American Economic Review* 87 (1997):847-62.

Berg, Joyce, Robert Forsythe, and Thomas A. Rietz, "What Makes Markets Predict Well? Evidence from the Iowa Electric Markets." *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten*, W. Albers, W. Güth, P. Hammerstein, B. Moldovanu, and E. Van Damme (eds.). New York: Springer (1996):444-63.

Camerer, Colin F. and Keith Weigelt, "Information Mirages in Experimental Asset Markets." *Journal of Business* 64 (1991):463-93.

Dyer, Douglas, John H. Kagel, and Dan Levin, "A Comparison of Naive and Experienced Bidders in Common Value Offer Auctions: A Laboratory Analysis." *Economic Journal* 99 (1989):108-15.

Forsythe, Robert and Russell Lundholm, "Information Aggregation in an Experimental Market." *Econometrica* 58 (1990):309-47.

_______, Forrest Nelson, George Neumann, and Jack Wright, "Anatomy of an Experimental Political Stock Market." *American Economic Review* 82 (1992):1142-61.

________, Thomas A. Rietz, and T. W. Ross, "Wishes, Expectations and Actions: A Survey on Price Formation in Election Stock Markets." *Journal of Economic Behavior and Organization* 39 (1999):83-110.

Hayek, Friedrich A., *Individualism and Economic Order*, University of Chicago Press, 1948.

Hung, Angela and Charles R. Plott, “Information Cascades: Replication and Extension to Majority Rule and Conformity-Rewarding Institutions” *American Economic Review*, 91, No.5 (2001), pp.1508-20.

Kagel, John H. and Dan Levin, "The Winner's Curse and Public Information in Common Value Auctions." *American Economic Review* 76 (1986):894-920.

King, Ronald R, Vernon L. Smith, Arlington W. Williams, and Mark V. Van Boening, "The Robustness of Bubbles and Crashes in Experimental Stock Markets." *Nonlinear Dynamics and Evolutionary Economics*, I. Prigogine, R. Day, and P. Chen (eds.). Oxford University Press (1993)

Lei, Vivian, Charles N. Noussair, and Charles R. Plott, "Non-Speculative Bubbles in Experimental Asset Markets: Lack of Common Knowledge of Rationality vs. Actual Irrationality." *Econometrica* 69, no. 4 (July 2001):831-59.

Lind, Barry and Charles R. Plott, "The Winner's Curse: Experiments with Buyers and with Sellers." *American Economic Review* 81, 1 (1991):335-46.

Charles R. Plott, “Markets as Information Gathering Tools”, *Southern Economic Journal*, 67 (1), 2000:2-15.

_____________, and Shyam Sunder, "Efficiency of Experimental Security Markets with Insider Information: An Application of Rational Expectations Models." *Journal of Political Economy* 90 (1982):663-98.

_____________________________, "Rational Expectations and the Aggregation of Diverse Information in a Laboratory Security Markets." *Econometrica* 56 (1988):1085-118.

_____________, Jörgen Wit, and Winston C. Yang, "Parimutuel Betting Markets as Information Aggregation Devices: Experimental Results. *Economic Theory*. Forthcoming.

Porter, David P. and Vernon L. Smith, "Stock Market Bubbles in the Laboratory." *Applied Mathematical Finance* 1 (1994):111-27.

Vernon L. Smith, Gerry L. Suchanek, and Arlington W. Williams, "Bubbles, Crashes and Endogenous Expectations in Experimental Spot Asset Markets." *Econometrica* 56 (1988):1119-51.

Sunder, Shyam, "Market for Information: Experimental Evidence." *Econometrica* 60 (1992):667-95.

________, "Experimental Asset Markets: A Survey." *Handbook of Experimental Economics*, John H. Kagel and Alvin E. Roth (eds.). Princeton University Press (1995).