Using Software Agents to Supplement Tests Conducted by Human Subjects

Hyungna Oh* and Timothy D. Mount**

Abstract

The objective of this paper is to test whether or not software agents can match the observed behavior of human subjects in laboratory tests of markets. For this purpose, one set of tests uses four software agents and two human subjects to represent six suppliers in three different market situations: no forward contracts, fixed price forward contracts, and renewable forward contracts. An identical set of tests is also conducted using software agents to represent all of the suppliers. The results show that software agents were able to replicate the behavior of human subjects effectively in the experiments. This indicates that software agents can be used effectively in testing electricity auctions, doing additional sensitivity tests, and supplementing results obtained using human subjects.

Keyword: Agent-based economics, electricity auction, experimental economics
JEL: C73, L94, D43.

1. Introduction

Restructured electricity markets have exhibited unsatisfactory results, most notably in California. Since electricity is a central component of modern economies, market operators and regulatory agencies continually introduce new types of market structures to obtain a more reliable electricity market. Recent introductions include a micro-grid, a capacity market, long-term contracts, demand-side participation, financial transmission rights, and customer's choice of retail services. More recently, smart electricity meters and real-time pricing have also been considered to improve efficiency and mitigate wholesalers' market power. Furthermore, deregulation and the unbundling of generation, transmission, and distribution functions provide many choices for a supplier, such as vertical integration, merging with other firms, entering into the new market, or divesting from the market. This variety of choices for generating firms, customers, and market operators implies that electricity markets are not fixed, but continue to change.

This type of evolving market requires suitable modeling tools that can be used to test the new market structures and new market rules before they are applied to real markets. Experimental economics has been used to test how wholesalers and consumers change their behavior when market conditions and rules are changed, and how spot prices are affected. However, there are some important restrictions on the design of an experiment. Viable results must be obtained using a relatively small number of suppliers and a relatively small number of trading periods. For example, the standard market test using PowerWeb at Cornell involves only six firms, and tests with more than 50 trading periods are rare. Agent-based simulation can be an alternative to laboratory tests using people. In an agent-based simulation, the role of human subjects in an experiment is taken by software agents. One advantage of using software agents, rather than people, to test markets is that it is practical to run a much more extensive range of tests.

A software agent with computational intelligence is a computer program representing an economic decision and performs its assigned task in a virtual environment. In order to perform the task efficiently, an agent has at least a perception and a decision function. The former receives new pieces of information and rearranges them to extract useful information, while the latter selects the best action to maximize its satisfaction. In this process, the algorithms rely on heuristic arguments and similarities to nature (Dawid, 1999).

Agent-based Computational Economics (hereafter, ACE[1]) is a branch of economics utilizing artificial

---

* Corresponding author: Assistant Professor, Department of Economics, 411 B&E, West Virginia University, Morgantown, WV 26506, Phone: 304-293-7870, Fax: 304-293-2233, Hyungna.Oh@mail.wvu.edu .
** Professor, Department of Applied Economics and Management, 215 Warren Hall, Cornell University, Ithaca, NY 14853-7801, Phone: 607-255-4512, TDM2@cornell.edu.

We are grateful to students who participated in power experiments at Cornell University. This work was supported in part by the US Department of Energy through the Consortium for Electric Reliability Technology

intelligence techniques in economic research. Advantages of these ACE approaches to electricity market research include the following: 1) an agent may generate different behaviors in the spot market even when market conditions are exactly the same as before. This time-variant strategy is the result of either learning effects or changes in market conditions, 2) the dynamic relation of suppliers' behavior to market price can be explained, since a software agent interacts with other agents, determines the market price and then reacts to the market price, 3) if new rules or market structures are specified, an ACE system can test how these alter the market price and the supplier's behavior, and 4) it can closely simulate real market conditions using diversified agents, with business goals, production conditions, risk attitudes and preferences, and accessibility to information different by agent.

There are various software agents designed for restructured electricity markets in the U.S. For example, EMCAS (Electricity market complex adaptive system) has been developed by Argonne National Laboratory and analyzes the technological and economical aspects of electric power systems. The learning used in EMCAS software agents is called "exploration-based learning". GenCo, a supplier agent in EMCAS, uses this learning and pursues its given goals (such as maximizing profits, maintaining the minimum market share, and avoiding regulatory intervention). Tesfatsion and her colleges at ISU developed AMES (Agent-based Modeling of Electricity Systems), and tested the market power, efficiency and reliability issues in wholesale electricity markets (2007). Talukdar et al at CMU simulated price spikes resulting from bidders' withholding behavior (2004) and showed how a cascading failure can occur, using autonomous adaptive software agents (2005). Cornell University combined autonomous software agents with experiments using human subjects in analyzing various market options and detecting market power (Oh and Thomas 2006; Oh et al 2005; Mount and Oh 2004).

In the UK, Bower and Bunn (2000) applied agent-based techniques to the new two-settlement market (Power Exchange and a Balancing Market), which included a discriminatory auction for the balancing market. The software agent in their simulation was designed to maximize profit conditionally on maintaining a minimum market share. As Bower and Oliveira (2001) recognized, they did not incorporate any learning effects. Bunn and Oliveira (2001) extended Bower and Bunn (2000) with reinforcement learning. Recently, Day and Bunn (2008) demonstrated that an ACE approach can be a very useful tool to explain continuously evolved strategies in a large scale and complex market.

The objective of this paper is to demonstrate that software agents can match the observed behavior of human subjects in laboratory tests of markets. For this purpose, one set of tests uses students to represent suppliers in an electricity auction with 1) no forward contracts (all dispatched capacity is paid the spot price), 2) permanent forward contracts (i.e. two suppliers hold a permanent forward contract, the same contract is held for all trading periods, and the price of this contract is independent of the spot prices), and 3) renewable forward contracts (i.e. a forward contract is renewed periodically and spot prices influence the forward price). An identical set of tests is also conducted using software agents (i.e. artificial intelligence) to represent all of the suppliers.

The analysis is based on simulations of a wholesale market for electricity run by an Independent System Operator (ISO). Suppliers submit offers into a central auction, and the ISO determines the optimum pattern of dispatch to minimize the cost of meeting load. A uniform price auction is used to determine the market price.

Our software agents (see Mount and Oh 2004) have a backward looking function to learn about the current market from the previous market outcomes. The adaptation involves updating an estimate of the residual demand curve faced by each firm, and this curve is used by the firm to determine the optimum set of offers to maximize expected profits in the next round of the auction. A noticeable result from our earlier work (Mount and Oh 2004) is that under the load uncertainty, agents replicate supply curves that are shaped like hockey sticks. This is exactly the type of behavior observed in deregulated electricity markets.

The results using software agents were encouraging. In the first set of tests, two students competed in each market with four software agents. By adjusting parameter values in the residual demand function, these four agents were designed as price-takers at the beginning. However, they can learn during the experiment, and evolve a Cournot strategy, a Bertrand strategy, or both. In almost all cases, the average earnings of the software agents were higher than the average earnings of the students. In the tests with all

---

Solutions (CERTS) and in part by the National Science Foundation Power Systems Engineering Research Center (PSERC).

[1] According to Tesfatsion's definition (2001), ACE is the computational study of economies modeled as evolving systems of autonomous interacting agents with learning capacities.

software agents, two software agents replaced the students. These two software agents were designed to consider that the electricity market can be imperfectly competitive. These two agents were designed by setting parameter values of their initial residual demand functions to give their shape a curvature. The average spot prices and average earnings with all software agents corresponded closely to the highest values obtained by the students. Our outcomes demonstrate that software agents can be used effectively to test electricity auctions, do additional sensitivity tests and supplement results obtained using humans.

The rest of this paper is organized as follows. Section 2 explains the experimental framework and the design on software agents. Section 3 compares the performance of human subjects and software agents, and tests whether or not software agents can be used to replicate the behavior of actual suppliers and people in laboratory tests. In Section 4, we conclude the paper.

2. The Experimental Framework

2.1. An Experimental Design

All of the experiments were conducted using *PowerWeb* (an interactive, distributed, Internet-based simulation platform developed by PSERC researchers at Cornell University) to test different electricity markets using human decision makers and/or computer agents. An Independent System Operator (ISO) determines the optimum dispatch of generators and the spot (nodal) prices paid to suppliers. The *PowerWeb* environment is designed to run unit commitment and optimal power flow routines to minimize the cost of meeting load subject to the physical constraints of an AC network. However, for our experiments, the network constraints are not binding, and, in each trading period, the same spot price is paid to all suppliers, using a uniform price auction (last accepted offer).

After each trading period, the ISO announces a forecast of the load in the next trading period. Load is completely price inelastic but it does vary from period to period. The forecasted load is generated randomly using a uniform distribution from 430 MW to 550 MW. The actual load is also generated randomly using a uniform distribution (Forecast ± 20 MW). The average load is 82% of the total installed capacity, which corresponds to realistic conditions in the summer when the load is relatively high.

For each trading period, each supplier submits offers to sell (or withhold) five blocks of capacity into the auction. A price cap (maximum offer allowed) of $100/MWh is enforced by the ISO (to keep the payments to participants in the test reasonably low). If the total capacity offered into the auction is less than the actual load, the ISO covers the shortfall by purchasing expensive imports from another market. However, when a shortfall occurs, the spot price is set by the highest offer and not by the price of imports.

Each supplier owns five blocks of generating capacity with capacities 50, 20, 10, 10, 10 (MW) and production costs 20, 40, 48, 50, 52 ($/MWh generated), respectively. In addition, there is a fixed standby cost of $5/MW to cover the opportunity cost of availability when a block is offered into the market. Withholding a block from the auction is the only way to avoid the standby cost for that block. There is also a fixed cost charged each period to cover capital costs ($1200/period, to make earnings roughly equal to profits in excess of competitive levels). These capacity and cost structures are the same for all six suppliers and they remain the same in all of the markets tested.

Three different markets, 1) no forward contracts (Test 1), 2) a permanent forward contract (Test 2), and 3) a renewable forward contract (Test 3), were tested during Fall semester, 2003 using 20 students majoring in applied economics or electrical engineering (the tests were part of a course on electricity markets). Each student represented a supplier in the market. In addition, some of the suppliers were represented by computer agents. In Test 1, none of the six suppliers holds a forward contract, but the two vertically integrated firms (agents) must meet one sixth of the load at a predetermined price. In Test 2, regulations require that each student must hold a forward contract for half of her capacity, and has already signed a contract for 50MW (the first block of capacity) at a fixed price of $60/MWh. These contracts are in place for all periods. Hence, the objective is to maximize the profits from selling the remaining four blocks of capacity (the first block is submitted automatically). In all other respects, the conditions are the same as in Test 1.

In Test 3, each student has to renew a 10-period forward contract of 50MW every 10th period. The forward price is given in (2) with λ=0.25 and a random residual added. The value of λ is not a priori information for the suppliers. The computer agents are designed to estimate λ based on the previous spot and forward prices. Simulation results show that the agents' estimates of λ are accurate after 3 periods. In

Test 2 and Test 3, the students were paid for the forward contracts as well as for the earnings in the spot market. In Test 3, the first contract price was set at \$60/MWh for periods 1 to 10, and this contract was renewed in periods 10 and 20. The students' earnings were computed to reflect the forward prices in the two new contracts and not the initial contract (the actual revenue from the first two contracts was augmented by $50 \cdot 10 \cdot (P_{20}^{F} - 60)$. The reason for doing this was to provide the students with the same incentive to increase the forward price of a new contract throughout the test.

For each test, there are 10 sessions with six suppliers in each session: two students, two VIF agents and two IPT agents. The lowest profit session in each test is excluded so that results correspond more closely to the behavior of professional traders in real markets.

The students in the tests represented experienced traders, and they received an initial briefing about how suppliers behave in the PJM market. Hence, the students understood the rationale for speculating, and why hockey-stick offers cause price spikes. Test 1 (no contract) consisted of 25 trading periods, and the next two tests, conducted a week later, consisted of 20 trading periods each. Each student was paid in direct proportion to her cumulative earnings and told that the objective of the tests was to earn as much money as possible. The initial trading periods were treated as learning periods for developing an offer strategy, and the average results from the last 10 periods in each test were used in the analysis.

In the first set of our experiments, two students compete with four agents. Two of these agents represent Vertically Integrated Firms (VIF) that have to meet a fixed proportion of load and are paid a regulated price (=\$60/MWh) for this load. These firms have less incentive to speculate than the others. The other two agents are "Initial Price Takers (IPT)" (believe that price spikes can not occur). However, initial price takers can learn to speculate if high prices do occur, and, in this sense, these two agents reinforce the behavior of the students. If the students do not speculate, none of the computer agents speculate, but if the students speculate, the agents learn to speculate and make the market easier to exploit.

An identical set of tests was also conducted using computer agents to represent all six suppliers by replacing the students with two "latent speculators (LS)" (believe price spikes can occur). These agents are more likely to speculate than initial price takers, and initial price takers evolve into latent speculators if there are high prices. The primary objective of the tests with all agents was to determine how well the computer agents can replicate the typical offer behavior of the students.

2.2. The Design on Software Agents

In our experiment, the task of each software agent is to compete with other software agents and human subjects and maximize its own profit. Like human subjects, each software agent owns five generating blocks, submits offers to the uniform price auction, and observes market outcomes when the market is cleared. All software agents determine their offers synchronously and independently. This reflects market rules that prevent one firm from communicating with another on offer strategies. We assume that each software agent considers that the electricity auction is continuously operated in this way in order to exclude the possibility of atypical offers (i.e., terminal offers) that a software agent may make at the end of the simulation period.

Software agents in this study anticipate forthcoming market conditions using the residual demand function. We specify a residual demand curve as an inverse function of the "excess" capacity offered into the auction (i.e. the available capacity that is offered but is not dispatched).

$$
\begin{aligned}
(1)\quad \mathrm{P} &= 1/(a_t + b_t(OC_t + q - \hat{Q}_t)) \\
&= 1/(a_t + b_t OC_t - b_t(\hat{Q}_t - q)) \\
&= 1/(\alpha_t - \beta_t(\hat{Q}_t - q)/IC)
\end{aligned}
$$

where P is the market price,

$\hat{Q}_t$ is the forecasted system load,

OC_t is the offered capacity from other firms,

IC is the installed capacity of other firms,

$q < q_{max}$ is the own capacity dispatched, and
$a_t > 0$ and $b_t > 0$ are the subjective parameter values of the firm.

The re-parameterization to $\alpha_t = a_t + b_t OC_t$ is convenient because OC_t is unobserved, and this avoids the computational problems of getting $a_t < 0$ when updating (b_t is also used in the updating process, but β_t is specified here because the values are easier to interpret). $P_L = 1/\alpha_t$ corresponds to the low market price if the firm could undercut the offers of all other firms and cover all of the load (i.e. $q = \hat{Q}_t$). Clearly, the firm's own installed capacity, $q_{\max}$, is the maximum that can actually be offered into the auction by a firm. For the other parameter ($\beta_t = b_t IC$), $P_H = 1/(\alpha_t - \beta_t)$ corresponds to the highest possible price in the market when $q = \hat{Q}_t - IC$ (i.e. the price for the first unit of capacity dispatched in the market).

This form of residual demand allows for a wide range of market behavior from competitive to the type of speculation implied by "hockey stick" supply curves (see Oh (2003)). In a truly competitive market, $P_L = P_H$ and $\beta_t = 0$. When $0 < \beta_t < \alpha_t$, $P_H > P_L$ and the firm believes that it has some market power. As $\beta_t \to \alpha_t$, P_H increases, and values greater than the price cap in the market can be interpreted as other firms withholding capacity from the auction. This type of withholding can be sufficiently large to make the firm "pivotal" (i.e. essential for meeting the load when $OC_t < \hat{Q}_t$). The restriction $0 < \beta_t \, \hat{Q}_t / IC < \alpha_t$ ensures that prices are positive and finite for $0 \le q \le q_{max}$, which is the relevant range of quantity offers for the firm ($\beta \, \hat{Q}_t / IC > \alpha_t$ makes the firm pivotal).

This linkage between parameter values and suppliers' behavior provides information on the design of software agents. For example, we set initial values of α_0 and β_0 in order that IPT agents' residual demand curve is almost flat and their expected market price is close to the competitive price for a given load: $\beta_0 = 0$ and $1/\alpha_0 = [P_0^C \mid load]$. For LS agents, initial value of α_0 is the same as IPT's but their initial value of β_0 is set to match $1/(\alpha_0 - \beta_0) = [P_0^H \mid load]$, where P_0^H is an IPT agent's expected market price for a given load. We can compute cost-based prices and determine the range of $P_0^C \in [45{,}55]$. The range of P_0^H is selected, $P_0^H \in [65{,}100]$, to cover 30 percent or above (up to price cap) mark-up pricing behavior. Then, a set of randomly drawn values is assigned for each software agent, and set initial values of α_0 and β_0 are used.

Two parameters, α_t and β_t, in the residual demand function are time-varying and revised whenever new information is available. New information is embodied in the price prediction error in equation (1). In each round, software agents adjust their price prediction in the previous round by applying the actual load in equation (1), compute the price prediction error, and adjust the two parameters. Therefore, changes in the parameters are proportional to the size of this prediction error. By applying load forecast for the current round, software agents evaluate market conditions and investigate whether or not they can be better off playing non-competitive strategies. This learning process can be termed a Kalman adaptive algorithm.

The main advantage of our approach is that the behavior of each firm agent can be evaluated directly using conventional economic criteria such as a supply function equilibrium (Klemperer and Meyer 1989). However, unlike human subjects, our software agents are incapable of learning about the structure of the market by employing complex counterfactual scenarios or deep introspection, and they rely on simple adaptive learning using a Kalman filter.

Once the residual demand function is updated, the software agent then determines the optimal offer for each block of capacity to maximize expected profits. A numerical search is used to determine the set of optimal offers. A numerical search consists of 6 steps: 1) generate a series of random numbers distributed around the load forecast, 2) draw a level of load (L_j) from the distribution and compute the residual demand function with it, 3) for a set of possible offer prices $P_i \in [0{,}100]$, compute profits from the first block (base block) to the block currently considered for L_j, 4) considering load uncertainty, compute the expected profits of each offer price, 5) the optimum offer for each block is determined to maximize the expected profit, and 6) then, if the expected profit of the optimum offer cannot increase the total expected profit, the block is withheld from the market to avoid paying stand-by costs.

In our earlier work (Mount and Oh 2004), we demonstrated that, in the presence of load forecasting errors, our design of software agents can replicate the observed supply curves, which are shaped like hockey sticks. Figure 1 shows the simulated offer curve with a price cap of $1,000/MWh in PJM. This is exactly the type of behavior observed in deregulated electricity markets.

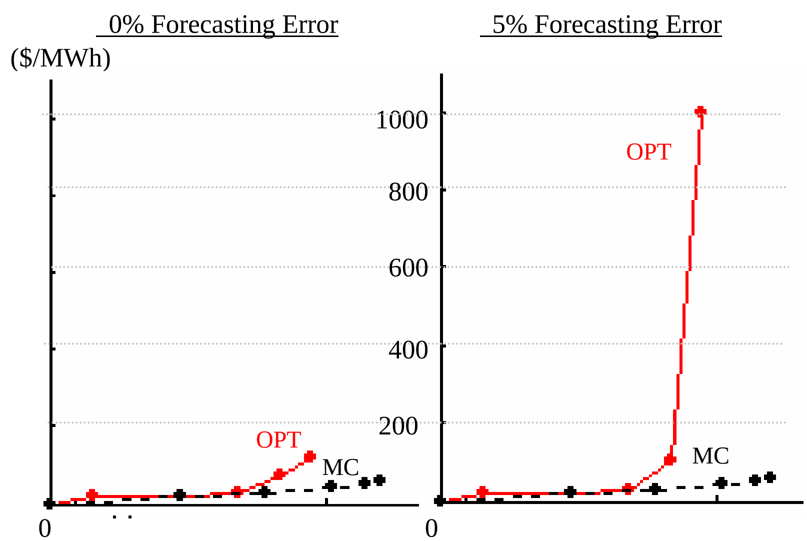

Figure 1. The Effect of Different Load Forecasting Errors (0% and 5%) on the Optimum Offers (OPT) of a Firm (MC is the Marginal Cost) (Mount and Oh 2004)

3. The Performance of Computer Agents

In the first part of this section, we will demonstrate that software agents can replicate the observed offer behavior of human subjects (students) and actual generating firms in the market. In the second part, the performance of the computer agents in the experiments is evaluated and compared to the performance of the two IPT agents and the two students, using the average earnings in the three tests. Then, in the third part of this section, we will compare the results of experiments using all agents with those of experiments using four agents and two human subjects.

3.1. Replication of the observed behavior of actual suppliers and human subjects in laboratory tests

Figure 2 gives one example of how students and software agents behave in the auction. This example uses the session of the highest earnings for Test 1 (session 6). The 9th (top) and 24th (bottom) periods are selected to show the learning effects in offer behavior of software agents. Note that forecasted load and actual load were similar in these two periods. In earlier periods (including the 2nd and 3rd), students submitted the maximum offers ($100MWh) for their 4th and 5th blocks and, as a result, students and software agents experienced very high market clearing prices ($100MWh). In the period of 9, IPT agents submitted 80 percent of their installed capacity, similar to the proportion of forecasted load to the total installed capacity in the market (600MWh), and their offer prices were above marginal costs. These offer behaviors were less competitive than their initial behaviors, but still more competitive than those of the students. However, as the auction was repeated, IPT agents evolved offer behaviors similar to students'. In the 24th period, IPT agents submitted the maximum offer prices ($100MWh) and generated a supply curve shaped like a hockey stick, as the students did. This example demonstrates how IPT agents learned from market outcomes, developed their strategies and, finally, demonstrated offer behavior similar to that of students in the laboratory tests, and actual suppliers in the market.

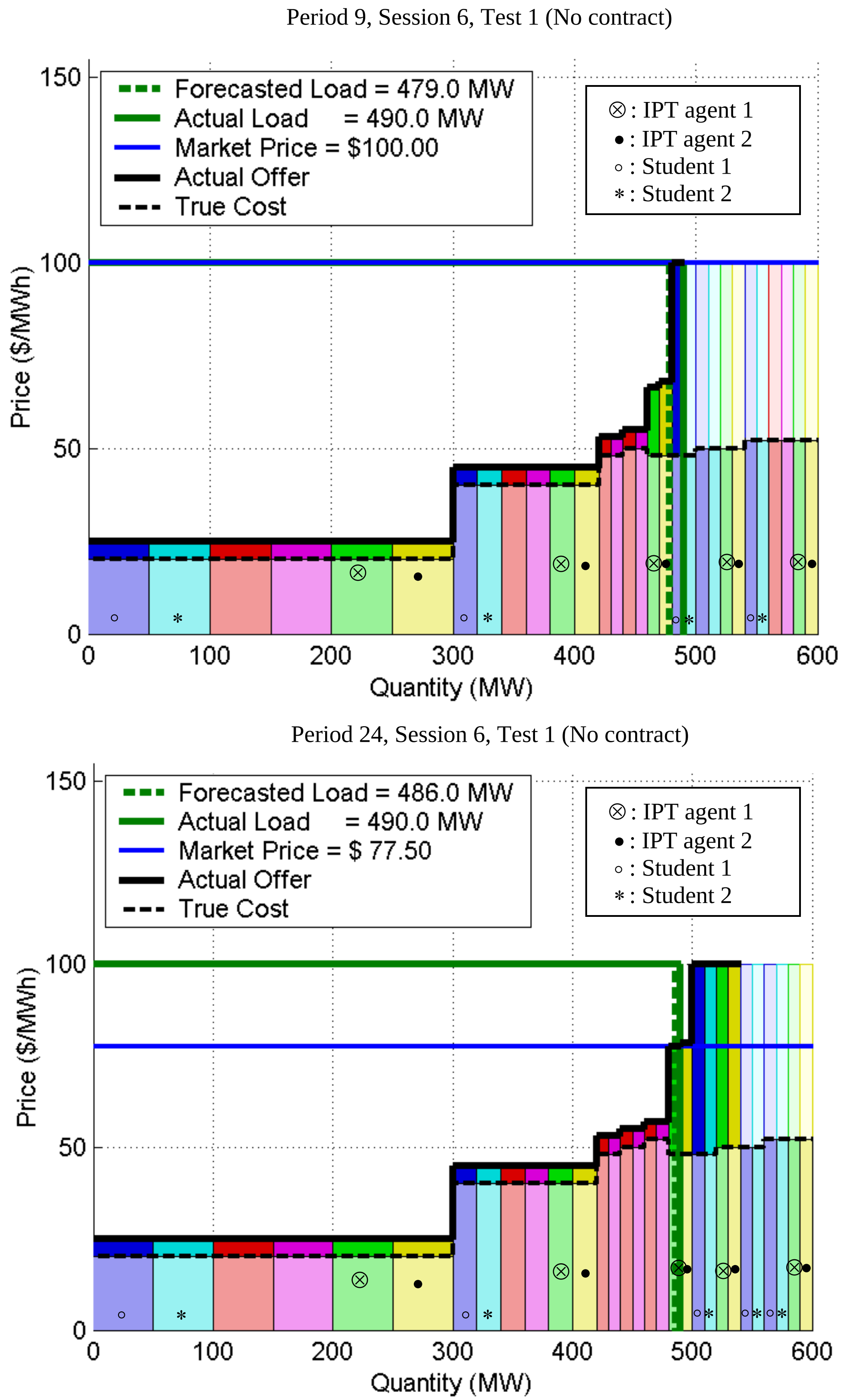

Figure 2. Learning Effects: Offers submitted into the auction in periods of 4 and 24 (faded areas represent capacity blocks withheld from the market)

3.2. Human subjects (Students) versus IPT agents

Table 1 summarizes the main results of the first set of experiment in a compact form. The average earnings of the IPT agents in different sessions were higher than the earnings of the students in all three tests, and significantly higher in Tests 2 and 3. The results for individual sessions are also depicted in Figure 3. A circle above the 45 degree line implies that the two IPT agents in a specific session earned more than the two students. This was true for most sessions in all three tests. For the nine sessions in a test, the earnings of the IPT agents were lower in 3 sessions for Test 1, but higher in every session for Tests 2 and 3.

Table 1. Average Earnings ($ in periods 11 to 20)

Variable	Students	VIF-Agents	IPT-Agents
Test 1	13,019	10,845	13,170
	(3,234)	(633)	(2,701)
Test 2	10,230	11,295	17,031
	(2,703)	(894)	(4,713)
Test 3	18,625	12,210	21,494
	(3,532)	(866)	(4,130)

* The standard deviation is given in parentheses

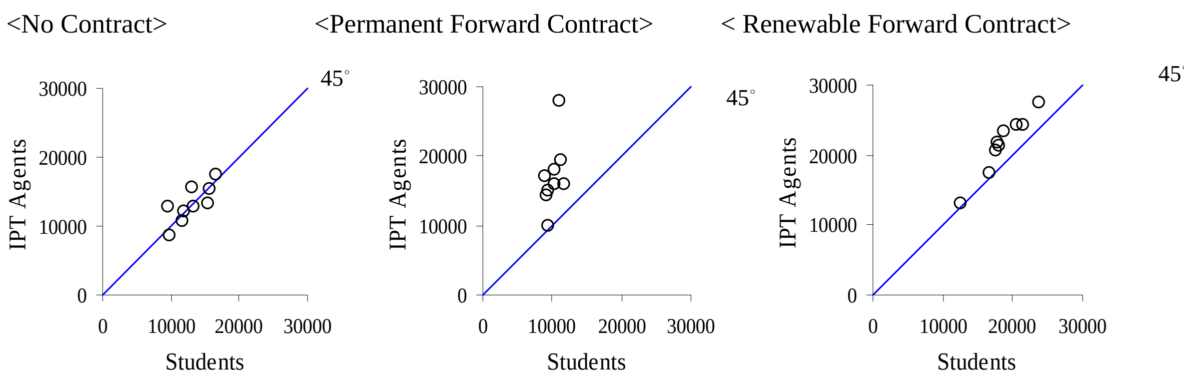

Figure 3. Comparison of Average Earnings

These results are particularly important for Test 1. In contrast to Tests 2 and 3, the earnings of the students and the IPT agents were similar in Test 1. Note that these were average earnings during 16-25 periods of Test 1, after a learning period of 15 periods. When the average earnings during the first 15 periods of Test 1 were compared, the IPT agents earned more than the students in 8 of the 9 sessions. Considering that all firms had the same cost structure for their generators, and that the students and the IPT agents faced identical market conditions, the IPT agents did very well compared to the students.

In Test 2, the IPT agents had a distinct advantage because the students held a contract for 50MW at a predetermined price. The objectives for the IPT agents and the students in Test 3 were different, but there was no clear advantage for the IPT agents. The results demonstrate that the IPT agents are i) just as good as the students in a simple market (Test 1), and ii) are able to exploit market conditions effectively (Test 2 and Test 3) when given the opportunity. This provides some preliminary evidence that computer agents can be used effectively to test the performance of simple electricity markets.

It is also clear from Figure 3 that the earnings of students and IPT agents are positively correlated. When one pair of students is successful in raising the market price, the IPT agents learn to speculate and reinforce the students' behavior. As a result, all firms get higher earnings. The IPT agents can also exploit unusual situations effectively. For example, in Session 2_9 of Test 2, one student sold only the contracted 50MW and withheld everything else from the auction, and the other student submitted three of the four non-contracted blocks (40MW in addition to the 50MW contracted) at very high offer prices. As a result, the spot prices were high. These high prices persisted because the two students did not change their behavior. Under these circumstances, the IPT agents earned more than 2.5 times as much as the students. The IPT agents withheld less capacity than the first student and submitted lower price offers than the second student. It was not necessary for the IPT agents to speculate, because the students were speculating so aggressively. This is exactly the type of strategy that was followed by Eastern in the UK market during the 1990's, when high market prices were set on a predictable basis by two other firms.

Meanwhile, the earnings of the VIF agents were lower than those of the students in Tests 1 and 3, but not in Test 2, when the students had permanent contracts. This is expected, since a significant proportion of VIF agents' capacity is committed to be sold at a fixed price ($60/MWh), which is a lot lower than market clearing prices ($75/MWh or above) .

3.3. Experiment with Human Subjects Versus Experiment with All Software Agents

When the three tests were repeated using computer agents to replace the two students, the average earnings of firms and the average market prices were higher than the corresponding values in all cases. For these tests, the students' firms were represented by Latent Speculators (LS). LS agents are more likely to speculate than IPT agents, but, when high prices occur, the IPT agents adapt to the new market conditions and evolve into LS agents.

The average earnings by the type of firm are summarized for the all-agent tests in Table 2. Percentage changes from the corresponding values in Table 1 are also shown, and, in eight out of the nine cases, these changes are positive. The small negative change for the VIF agents in Test 2 is the only exception. The positive changes for Test 1 (no contract) and Test 3 (renewable contract) are very large (ranging from 13% to 70%). It is only when the LS agents have a permanent contract in Test 2 that the changes are relatively small (ranging from -1% to 21%). A comparison of the average earnings of the LS agents in Table 2 to the corresponding session values for the students in Table 1 shows that the values for the LS agents fall in the ranges observed for the students. For Tests 1 and 3, the earnings of the LS agents are similar to the highest earnings of the students, but, for Test 2, they are only slightly above the median value. The general conclusion is that the LS agents were able to exploit market power effectively when the opportunity arose in Tests 1 and 3. However, more students were able to get higher earnings than the LS agents in Test 2 when it was relatively difficult to exploit market power.

Table 2. Average Earnings ($/MWh)* for All-Agent Tests

Variable	LS-Agents	VIF-Agents	IPT-Agents
Test 1	18,154	12,297	20,892
	(+39%)	(+13%)	(+59%)
Test 2	12,375	11,200	18,235
	(+21%)	(-1%)	(+7%)
Test 3	26,257	16,080	36,512
	(+41%)	(+32%)	(+70%)

* The percentage change from the corresponding value in Table 1 is given in parentheses

Using a Chow Type II test, it is possible to test whether or not the 18 new observations obtained from the all-agent tests deviated from the sample of 162 observations using students. A regression model is specified to make it easy to test the hypothesis as follows:

$$(2)\quad y_{ijk} = \mu + \sum_{i=2}^{3} \alpha_i M_i + \sum_{i=1}^{3}\sum_{j=2}^{3} \beta_{ij} F_{ij} + \sum_{i=1}^{3}\sum_{k=1}^{K-1} \gamma_{ik} S_{ik} + e_{ijk}$$

where y_{ijk} = log earnings for periods 11 to 20 for firm type *j* in session *k* of market *i*.

M_i = 1 for Test *i*, 0 otherwise

F_{ij} = 1 for Test *i* and Firm *j* , 0 otherwise

j = 1 for a student, 2 for a VIF agent and 3 for IPT agent

S_{ik} = 1 for Test *i* and Session *k*, -1 for Test *i* and the last session of each test, 0 otherwise

Table 3. Estimation Results of Equation (2): Dependent variable = average earnings

	Model 1			Model 2 (agents' earnings = average)			Model 3 (agents'earnings=the highest)		
Variable (Parameter)	Estim.	t-value	Pr>|t|	Estim.	t-value	Pr>|t|	Estim.	t-value	Pr>|t|
Intercept (μ)	9.441	252.33	<.0001	9.477	230.57	<.0001	9.451	265.12	<.0001
Market2 (α_2)	-0.242	-4.57	<.0001	-0.256	-4.40	<.0001	-0.247	-4.90	<.0001
Market3 (α_3)	0.374	7.07	<.0001	0.373	6.42	<.0001	0.370	7.34	<.0001
Market1VIF (β_{12})	-0.151	-2.85	0.0051	-0.175	-3.01	0.0031	-0.175	-3.49	0.0006
Market1IPT(β_{13})	0.024	0.46	0.6445	0.036	0.62	0.5357	0.036	0.72	0.4717
Market2VIF (β_{22})	0.130	2.46	0.0151	0.107	1.85	0.0670	0.107	2.15	0.0336
Market2IPT (β_{23})	0.510	9.64	<.0001	0.498	8.56	<.0001	0.498	9.96	<.0001
Market3VIF (β_{32})	-0.407	-7.69	<.0001	-0.415	-7.15	<.0001	-0.415	-8.31	<.0001
Market3IPT (β_{33})	0.141	2.66	0.0088	0.160	2.75	0.0068	0.160	3.19	0.0017
Session1_2 ($\gamma_{1,2}$)	0.013	0.22	0.8289	0.013	0.19	0.8520	0.007	0.11	0.9136
Session1_3 ($\gamma_{1,3}$)	0.086	1.41	0.1597	0.086	1.22	0.2240	0.080	1.32	0.1905
Session1_4 ($\gamma_{1,4}$)	0.094	1.54	0.1253	0.094	1.33	0.1848	0.088	1.44	0.1506
Session1_5 ($\gamma_{1,5}$)	-0.096	-1.57	0.1183	-0.096	-1.36	0.1766	-0.103	-1.69	0.0925
Session1_6 ($\gamma_{1,6}$)*	0.206	3.36	0.0010	0.206	2.90	0.0042	0.259	5.70	<.0001
Session1_7 ($\gamma_{1,7}$)	-0.050	-0.82	0.4149	-0.050	-0.71	0.4811	-0.057	-0.93	0.3521
Session1_8 ($\gamma_{1,8}$)	0.138	2.26	0.0256	0.138	1.95	0.0530	0.131	2.17	0.0319
Session1_9 ($\gamma_{1,9}$)	-0.130	-2.12	0.0358	-0.130	-1.83	0.0690	-0.136	-2.25	0.0262
Session2_2 ($\gamma_{2,2}$)	0.099	1.62	0.1068	0.099	1.40	0.1628	0.106	1.74	0.0838
Session2_3 ($\gamma_{2,3}$)	-0.232	-3.80	0.0002	-0.232	-3.28	0.0013	-0.226	-3.73	0.0003
Session2_4 ($\gamma_{2,4}$)	0.056	0.92	0.3596	0.056	0.79	0.4285	0.063	1.03	0.3044
Session2_5 ($\gamma_{2,5}$)	-0.034	-0.56	0.5798	-0.034	-0.48	0.6324	-0.028	-0.45	0.6501
Session2_6 ($\gamma_{2,6}$)	0.032	0.52	0.6069	0.032	0.45	0.6567	0.038	0.62	0.5335
Session2_7 ($\gamma_{2,7}$)	-0.064	-1.05	0.2964	-0.064	-0.91	0.3668	-0.058	-0.95	0.3429
Session2_8 ($\gamma_{2,8}$)	-0.077	-1.26	0.2086	-0.077	-1.09	0.2770	-0.071	-1.17	0.2445
Session2_9 ($\gamma_{2,9}$)*	0.221	3.62	0.0004	0.221	3.13	0.0021	0.170	3.76	0.0002
Session3_2 ($\gamma_{3,2}$)*	0.208	3.40	0.0009	0.207	2.93	0.0039	0.291	6.42	<.0001
Session3_3 ($\gamma_{3,3}$)	0.055	0.91	0.3659	0.055	0.78	0.4346	0.045	0.74	0.4599
Session3_4($\gamma_{3,4}$)	-0.342	-5.59	<.0001	-0.342	-4.83	<.0001	-0.352	-5.80	<.0001
Session3_5 ($\gamma_{3,5}$)	-0.023	-0.37	0.7091	-0.023	-0.32	0.7472	-0.033	-0.55	0.5834
Session3_6($\gamma_{3,6}$)	-0.105	-1.71	0.0891	-0.105	-1.48	0.1413	-0.115	-1.90	0.0596
Session3_7($\gamma_{3,7}$)	0.098	1.61	0.1098	0.098	1.39	0.1666	0.088	1.45	0.1495
Session3_8($\gamma_{3,8}$)	-0.012	-0.19	0.8492	-0.012	-0.16	0.8695	-0.022	-0.36	0.7157
Session3_9($\gamma_{3,9}$)	0.119	1.95	0.0537	0.119	1.68	0.0949	0.108	1.79	0.0758
Nobs	162 (=1st set of samples)			180 (=two sets of samples)			180(=two sets of samples)		
Sum of Squ.Errors	3.2503			4.9673			3.6759		

* sessions with the highest earnings obtained by the students for each market

The first null hypothesis assumed that the earnings of the all-agent firms were equal to the **average** earnings of the students (i.e. by setting the session effects for the all-agent tests to zero). The parameters in model (2) were estimated using the pooled data set of 180 observations. Estimation outcomes from the first sample of 162 observations and the second sample of 180 observations using are summarized in Table 3. The computed F statistic (3.82) is large and the null hypothesis is rejected (the critical value for an $F_{(18,130)}$ is 1.70 at the 5% level of significance). This implies that the earnings of the all-agent firms were statistically different from the average earnings in the tests using students.

The second null hypothesis assumed that the earnings of the all-agent firms were equal to the sessions with the **highest** earnings obtained by the students (i.e. by selecting the sessions with the largest positive session-coefficients for each market in Table 3 (Sessions 6, 9 and 2 for Markets 1, 2 and 3, respectively)). In this case, the computed F statistic (0.95) is small and it supports the null hypothesis (note that the critical value of 1.70 is still valid). In other words, the earnings for the all-agent firms were statistically equivalent to the sessions with the highest earnings obtained by the students.

Average prices in the all-agent tests are summarized in Table 4, and percentage changes from the corresponding average prices from the first sample are also reported. The price changes in the three tests are all positive and equal to 14%, 6% and 28%, respectively. With a renewable contract in Test 3, the average price in the all-agent test ($95.7/MWh) is substantially higher than the average price ($74.9/MWh) obtained by the students (see Oh and Mount 2005 for this issue).

Overall, the results from the all-agent tests are encouraging, and show that computer agents do provide a valid means of evaluating the performance of electricity markets. The computer agents were able to match the earnings of the best students. It will be interesting to find out in the future whether this also proves true for more complicated market structures, such as joint markets for energy and ancillary services.

Table 4. Average Market Prices ($/MWh) for All-Agent Tests

	4 Agents + 2 Students		All-Agents Test*	
Experiment	Average Spot Price	Average Forward Price	Average Spot Price	Average Forward Price
Cost-based	53.8	-	53.8 (0%)	-
Test 1: No Contracts	67.0	-	75.9 (+14%)	-
Test 2: Permanent Contracts	70.3	60.0	73.5 (+6%)	60.0 (0%)
Test 3: Renewable Contracts	76.4	72.4	95.7 (+28%)	92.9 (+30%)

*The percentage change from the corresponding value in the first set of tests is given in parentheses

4. Summary and Conclusions

The primary objective of this paper is to investigate how well computer agents can replicate the behavior of human subjects in tests of electricity auctions. Using PowerWeb to simulate the operation of a uniform price auction run by an ISO, four computer agents and two human subjects (graduate students) represent six supply firms in three different market situations. In each case, the patterns of load are exogenous and there are 20 trading periods (25 for Test 1). Three market structures are tested: 1) no forward contract (all dispatched capacity is paid the spot price), 2) the two students hold a permanent forward contract (the contract price is fixed), and 3) the two students hold a renewable forward contract (the current spot price influences the forward price used to renew the contract). In a second experiment, the three tests are repeated with two additional computer agents replacing the students.

The results for the computer agents were reassuring. Using the maximization of expected profits as the objective criterion for submitting offers by an agent, it is possible to modify the general form of a computer agent to represent different types of firm, such as a vertically integrated firm, and to deal with the different tests of forward contracts. In Test 1, with two students and four computer agents, the IPT agents and the students faced identical cost conditions. In 6 out of the 9 sessions in Test 1, the IPT agents had higher average earnings than the students. In the all-agent tests, the earnings of the LS agents that replaced the students were higher than the corresponding average earnings of the students in all three tests. In Tests 1 and 3, the earnings of the LS agents were similar to the highest earnings obtained by the students.

Our conclusion about the objective of the paper is that computer agents can replicate the behavior of students in an electricity auction effectively. In fact, the agents were also able to exploit unusual situations by, for example, behaving as free riders when the students in a session speculated aggressively. These results suggest that it is appropriate to do additional sensitivity tests using all computer agents. This is a promising line of research that is essential for developing realistic simulation models of deregulated electricity markets. Relying exclusively on human subjects to test different market structures will, due to the practical difficulty of recruiting enough people, limit the scope of the tests. Computer agents that can replicate realistic behavior can be used to extend the range and number of tests conducted with human subjects. This capability has tremendous potential for identifying potential flaws in market designs and finding effective ways to improve the performance of electricity markets before a specific market design is imposed on the public.

References

Bower, J. and Bunn, D., 2000, A Model-based comparison of pool and Bilateral Market Mechanisms for Electricity Trading, the Energy Journal, 21(3), 1-29.

Bower, J. and Bunn, D., 2001, Experimental analysis of the efficiency of uniform-price versus discriminatory auctions in the England and Wales electricity market, Journal of Economic Dynamics & Control, 25, 561-592.

Bunn, D. and Oliveira, F., 2001, Agent-based simulation: an application to the new electricity trading arrangements of England and Wales, IEEE Transactions on Evolutionary Computation, 5 (5), 493-503.

Bunn, D. and Day, C.J., 2008, Computational modeling of price formation in the electricity pool of England and Wales, Journal of Economic Dynamics & Control, forthcoming.

Dawid, H., 1999, Adaptive learning by generic algorithms. 2nd edition, New York: Springer.
Day, C.J. and Hobbs, B.F., 2002, Oligopolistic competition in power networks: A conjectured supply function approach, IEEE Trans. Power Systems, 17(3), 597-607.

Hines, P. and Talukdar, S., 2007, Controlling cascading failures with cooperative autonomous agents, International Journal of Critical Infrastures, 3(1-2), 192-220.

Klemperer, P.D., Meyer, M.A., 1989, Supply function equilibria in oligopoly under uncertainty, Econometrtica 57(6), 1243-1277.

Mount, T. and Oh, H., 2004, On the first price spike in summer, Proceedings of 37th Annual Hawaiian International Conference on System Science (HICSS), Big Island.

Oh,, H., 2003, Simulation methods for modeling offer behavior and spot prices in restructured markets for electricity, Doctoral thesis, Cornell University.

Oh, H. and Mount, T., 2005, Testing the effects of holding forward contracts on the behavior of suppliers in an electricity auction", Proceedings of 38th Annual Hawaiian International Conference on System Science (HICSS), Big Island.

Oh, H., Thomas, R.J., Leiseutre, Bernard C., and Mount, Timothy, 2005, A method for classifying offer strategies observed in an electricity market, *Journal of Decision Support System*, 449-460.

Oh, H. and Thomas, R.J., 2006, An agent-based optimal bidding function, Proceedings of 39th Annual Hawaiian International Conference on System Science (HICSS), Big Island.

Sun, J. and Tesfatsion, L., 2007, Dynamic testing of wholesale power market designs: an open-source agent-based framework, Computational Economics 30, 291-327.

Weiss, G., 1999, Multiagent systems: a modern approach to distributed artificial intelligent. The MIT Press.

Zhou, H., Tu, Z. and Talukdar, S., 2004, Wholesale electricity market failure and the new market design, Conferences in Computational Social and Organizational Science..