

An assessment of strategies for choosing between competitive marketplaces

T. Miller^{a,*}, J. Niu^b

^a*Department of Computer Science and Software Engineering, University of Melbourne, Parkville, 3010, Australia*

^b*CAISS, City College, CUNY, 137th Street and Convent Avenue, New York, NY 10031, USA*

Abstract

Traders that operate in markets with multiple competing marketplaces must often choose with which marketplace they will trade. These choices encourage marketplaces to seek competitive advantages against each other by adjusting various parameters, such as the price they charge, or how they match buyers and sellers. Traders can take advantage of this competition to improve utility. However, appropriate strategies must be used to decide with which marketplace a trader should shout. In this paper, we assess several different solutions to the problem of marketplace selection by running simulations of double auctions using the JCAT platform. The parameter spaces of these strategies are explored to find the best performing strategies. Results indicate that the softmax strategy is the most successful at maximising trader profit and global allocative efficiency in both adaptive and non-adaptive markets. The ϵ -decreasing strategy performs well in adaptive markets, while also showing greater stability in its parameter space than softmax. All marketplace selection strategies outperform the random marketplace selection strategy.

Keywords: Double auction, Mechanism design, CAT Game, Reinforcement Learning

1. Introduction

Due to the analytical complexity, most modern economic theory only considers isolated market institutions. However, in real markets, *marketplaces* compete against each other for buyers and/or sellers. This competition encourages marketplaces to improve their efficiency and reduce their prices to gain a larger market share, when compared to a monopoly.

*Corresponding author.

Email addresses: tmiller@unimelb.edu.au (T. Miller), jniu@gc.cuny.edu (J. Niu)

URL: <http://www.csse.unimelb.edu.au/~tmill/> (T. Miller),

<http://www.cs.gc.cuny.edu/~jniu/> (J. Niu)

Buyers and sellers can take advantage of competition between marketplaces by searching out the marketplaces that provide them with the highest reward. For example, a buyer may want to search out the lowest price for a given commodity. Having found a marketplace that attracts sellers willing to trade a particular commodity for less than other marketplaces do, the buyer will continue to buy from that marketplace.

While some of the information that is relevant to the traders is publicly available, such as price, other details may be private to the marketplace itself, such as the mechanism used to match buyers with sellers. To find its preferred marketplace, a trader must therefore explore the marketplaces a number of times to determine which marketplaces provide the best return for the trader. It can then exploit the feedback obtained from this sampling to improve its average future reward. Finding a balance between exploring and exploiting is known as the *exploration vs. exploitation trade-off*.

In automated trading systems, marketplace selection must be automated, therefore, algorithms are required to determine when to explore and when to exploit. Given a choice of N marketplaces with which a trader will interact many times, the trader must explore each of the N marketplaces several times to provide an estimate of its future reward from those marketplaces. Such a problem fits well into the framework of the well-known *N-armed bandit problem* (Sutton and Barto, 1998). The difference between the classical N-armed bandit problem and marketplace selection is that the payoff from different marketplaces are dynamic due to the fact that other traders in the market are solving the same problem at the same time and marketplaces may adapt their mechanisms over time.

Several solutions to the bandit problem have been proposed, however, it is not known which solution is the most effective for the marketplace selection problem. Finding efficient solutions to this problem has become more important in recent years as more stock and commodity trades are performed automatically, and as more off-market trading venues appear, giving traders more choice in marketplaces.

In this paper, we assess several of the most suitable algorithms for approaching the dynamic N-armed bandit problem, and we explore the parameter spaces of these algorithms to find the best solutions for the marketplace selection problem. This assessment is performed using the JCAT double auction simulation platform (Niu et al., 2008), presented in Section 3.

The algorithms assessed in this paper are: 1) the ϵ -first algorithm; 2) the ϵ -greedy algorithm; 3) the ϵ -decreasing algorithm; and 4) the *softmax* algorithm (Sutton and Barto, 1998). In addition, we assess a *random* choice algorithm as a baseline. These are discussed further in Section 4. We measure the profit that traders can make using each of these algorithms, as well as general market measures such as allocative efficiency, to explore whether intelligent marketplace selection algorithms can improve overall market performance on these aspects.

Two sets of experiments are run for each of these strategies: one in which the marketplaces attempt to adapt to the dynamics of the market by adjusting their fees across

trading days (*adaptive* marketplaces); and one in which the marketplace fees are static over all trading days (*non-adaptive* marketplaces). The selection strategies are *homogeneous*; that is, in each run, all traders employ the same strategy with the same parameters. The experimental setup is described in Section 5.

Our results, presented in Section 6, indicate that the softmax strategy obtains the highest average daily trader profit and average global allocate efficiency in both adaptive and non-adaptive markets. The ϵ -decreasing strategy performs well in adaptive markets, and shows greater stability in its parameter space than softmax. All marketplace selection strategies outperform the random marketplace selection strategy.

2. Related Work

Our work is aligned with Niu et al. (2007) and Cai et al. (2008). Niu et al. (2007) is the first of the kind to examine the competition between multiple marketplaces based on simulations using JCAT. This work showed that solution concepts to the N-armed bandit problem were effective in the scenario of marketplace selection and performed well even when trading strategies or other configurations vary. Our work considers a broader range of marketplace selection strategies and aims to explore the parameter spaces so as to find the most effective algorithms in different settings. Cai et al. (2008) used the JCAT platform to experiment with the economic effects that competition has in double auction markets. A result that they demonstrated is that having multiple competing marketplaces leads to a loss of global efficiency compared to a single marketplace containing the same traders. This is largely due to the trader distribution becoming fragmented. However allowing traders to migrate between marketplaces mitigates this loss of efficiency. This important result motivated our work in particular, as it demonstrated the importance of having effective marketplace selection strategies.

Sohn et al. (2009) discussed the effect of pricing policies on trader migration, and presented a pricing policy that attracts high-value intra-marginal traders — that is, traders that fall to the left of the equilibrium price of the market. Their results demonstrated that traditional adaptive traders cannot take advantage of this policy, but that traders that are aware of the policy can. The authors hypothesised that marketplace-specific trading strategies should be employed that learn the best strategy per marketplace, rather than an overall strategy.

In addition, the marketplace selection problem has been studied with formal approaches. For example, Rochet and Tirole (2003) and Shi et al. (2010) used game theory to analyse market performance with traders adopting a uniform marketplace selection. In contrast, our analysis is performed primarily from the perspective of the traders themselves — that is, which marketplace selection strategy should be used. Additionally, we use empirical evaluation to answer this question, as a game-theoretic analysis would not be suitable due to the complexity over the problem.

Marketplace selection or trading across multiple marketplaces have been studied

in other scenarios as well. Ladley and Bullock (2005) studied the market dynamics involving multiple marketplaces, but their work differed from ours on multiple aspects. First, they were concerned with the information available to traders. Traders in their analysis were each fixed to a certain location in a spatial network and could only trade with and receive information from their neighbours, while traders in our work, though spreading across multiple marketplaces in a similar way, have the capability of moving between these marketplaces based on their desire to maximise their profits. Second, Ladley and Bullock were concerned with how the different levels of traders' accessibility to market information affected the convergence of the whole market to the theoretical equilibrium, while our work focuses upon how the methods that guide the movement of traders perform differently from each other. Third, Ladley and Bullock constantly used a single classic continuous double auction mechanism in their experiments, while our work involves multiple marketplaces in direct competition that are each associated with an auction mechanism.

Our work also has similarities to that of Greenwald and Kephart (1999). In their work, shoppers choose between different merchants, and merchants set prices that depend on the prices set by other merchants. While shoppers and merchants are respectively analogous to traders and marketplaces in our work, the scenario we are considering is considerably more complex. Traders in our scenario learn over time using their profits as feedback in selecting marketplaces, while shoppers in Greenwald and Kephart's scenario either choose a merchant randomly or choose the merchant that offered the lowest price. The expected return of choosing a marketplace in our work is non-deterministic and hinges upon various factors that a trader has no way to know exactly about in advance or at all, e.g., other traders that go to the marketplace at the same time and the mechanism adopted by the marketplace. In contrast, shoppers in the work of Greenwald and Kephart know exactly about their utilities if they choose to buy from a particular merchant in the retail market. Indeed, the transaction prices are set by the merchants in this scenario, while in our case the prices are determined by the traders. As a result, when traders pick a marketplace in our scenario, they do not know for sure if they will even be able to trade, much less about the price at which goods will change hands.

Another piece of related work is Ganchev et al. (2010), which presented an algorithm for optimising the amount of trades to be placed over multiple competing dark pools. A dark pool is a type of exchange in which the volume of orders and the identification of the buyer and seller are not revealed. This allows traders to buy and sell large orders without revealing their trade to the rest of the market. Given a large set of trades, the trader must determine which proportion of trades to make in each pool at any time. If the number of trades submitted by a trader to a pool is too small, the pool is underutilised by that trader, and the trader cannot use the feedback to determine the maximum number of trades possible on that day. If the number of trades is too large, the trader can accurately determine the maximum number of trades, but some items will not be sold. The algorithm proposed by Ganchev et al. is a standard reinforcement learning algorithm based

on the R-MAX learning algorithm (Brafman and Tennenholtz, 2003). The algorithm is empirically tested on data from a large trading firm, and is compared to both a naive bandit algorithm, which appears to be a softmax algorithm, and a uniform allocation algorithm — that is, one that places an even number of orders on each exchange. The results demonstrated that the proposed algorithm outperformed both the bandit algorithm and the uniform algorithm, and also performed close to optimal. But different from our focus, they were concerned about how much liquidity each marketplace has and how to spread the request of trades in a way so as to maximise the volume of successful trades regardless of the profit that is made in doing so.

3. Background

To experiment with multiple marketplaces, we used JCAT (Niu et al., 2008), the software platform for the CAT Tournament (Niu et al., 2010), which allows marketplaces to run in parallel and traders move between them. Each entrant in the game provides a marketplace while traders are provided by the game organiser.

A CAT game lasts a certain number of days. Each entrant adopts various policies to regulate its marketplace, including a charging policy that decides the fees to impose on traders, and may adapt its policies over time so as to attract more traders and make more profit. A trader needs to choose a single marketplace on each day to trade with other traders that go to that same marketplace and may choose different marketplaces on different days.

Each trader is assigned a private value for the traded good. The private values and the number of goods to buy or sell determine the supply and demand of the market. The private values remain constant during a day, but may change from day to day. Each trader is also endowed with a trading strategy to decide how to make offers; e.g., the well-known strategy in the literature that is known as zero intelligence with constraint or ZI-C (Gode and Sunder, 1993); and a marketplace selection strategy to choose the marketplace to make offers in. The second strategy is our focus in this paper. These two tasks allow our traders to exhibit intelligence in two, orthogonal, ways.

4. Strategies for choosing marketplaces

This problem of marketplace selection can be formulated as a *dynamic* N-armed bandit problem (Sutton and Barto, 1998). The problem is dynamic for two reasons:

1. the strategies for the marketplaces can be adaptive, meaning that the underlying probability distributions can change over the course of a game; and
2. the rewards for marketplace selection are generated from distributions that are, in part, determined by the behaviour of other traders in the marketplace. For example, a buyer may choose a particular marketplace that has no sellers on that day, making it impossible for the buyer to receive a match.

Such a problem is different to other dynamic N-armed bandit problems, such as the *restless bandit* problem (Whittle, 1988), due to the fact that both the marketplaces and the other traders can affect the underlying probability distribution, which cannot be represented in a straightforward manner using a Markov decision process.

We assess four well-studied reinforcement learning algorithms for solving the N-armed bandit problem to determine which solution is the most suitable for choosing a marketplace, and what the parameters for these algorithms should be. In addition, we compare them to the baseline of a random choice algorithm. Each of these learning algorithms manages the exploration vs. exploitation problem using one or more parameters. The parameter space is explored as part of our study.

4.1. Decision-making algorithms

The marketplace selection problem can be solved using reinforcement learning, and is therefore split into two parts: the feedback, which determines how good the action was at achieving its goal; and the decision making, which determines which action the trader should choose next. In this section, we overview the decision making algorithms that we assess.

4.1.1. ϵ -greedy

The ϵ -greedy strategy selects the best action with a probability of $1 - \epsilon$, in which the best action is based on prior experience (exploitation), and ϵ is a parameter provided to the algorithm. For the remaining ϵ cases, the algorithm will randomly (with a uniform distribution) choose between all of the other available actions (exploration).

For example, if $\epsilon = 0.1$, the best action will be chosen approximately 90% of the time.

4.1.2. ϵ -first

The ϵ -first strategy chooses a random action each time for the first $\epsilon \cdot A$ actions (exploration), where A is the total number of times that an action must be chosen (the number of trading days in a CAT simulation). The best action is chosen for the remaining $(1 - \epsilon) \cdot A$ actions (exploitation).

For example, if $\epsilon = 0.1$ and $A = 1000$, a random action will be chosen, with uniform distribution, for the first 100 actions. Following this, the action that performed the best over those 100 trials will be chosen the remaining 900 times.

4.1.3. ϵ -decreasing

The ϵ -decreasing is similar to the ϵ -greedy strategy, except that the value of ϵ decreases over time. That is, the amount of exploration decreases as the agent learns more. This requires a second parameter, α , which specifies the rate of decay of ϵ , such that after each action, the value of ϵ becomes $\epsilon_0 \cdot \alpha$, in which ϵ_0 is the value of ϵ in the previous round, until a preset minimum is reached.

For example, if $\epsilon = 0.1$ and $\alpha = 0.9$, then after the first action, ϵ will be 0.09, meaning that there will be a 9% chance of exploration, and a 91% of exploitation. After the second action, ϵ will be 0.081.

4.1.4. *Softmax*

Softmax is a probability matching strategy, meaning that the probability of an action being chosen is dependent on the utility it has generated in previous rounds relative to all other available actions.

Our implemented softmax algorithm using a Gibbs (or Boltzman) distribution to select the next action:

$$\frac{e^{Q(a)/\tau}}{\sum_{b=1}^n e^{Q(b)/\tau}},$$

in which $Q(a)$ is the value of an action based on feedback from previous applications of that action, and τ is the *temperature*, a positive number that dictates how much of an influence the past data has on the decision. A high temperature causes the probabilities of actions to be closer to each other, while a low temperature causes them to be close to their $Q(a)$ values.

A second parameter, α , can be used to specify a rate of decay for τ , such that after each action, the value of τ becomes $\tau_0 \cdot \alpha$, in which τ_0 is the value of τ in the previous round, until a preset minimum is reached.

4.1.5. *Random*

The random strategy simply chooses between all marketplaces using a uniform distribution. No feedback is incorporated into the algorithm, and therefore, no exploitation phase is considered.

4.2. *Feedback*

The feedback mechanism employed is constant over all of the strategies considered. The algorithm used is an adapted version of the Widrow-Hoff learning algorithm (Widrow and Hoff, 1960), where the input regarding a certain marketplace is a weighted average of the previous profits the trader made in the marketplace, with the most recent days weighted heavier, and the output is the estimated profit that the trader can make on the current day if it chooses to trade in this marketplace. Given the potential dynamic behaviours of both traders and marketplaces, the use of the weighted average of daily profits helps to update the knowledge the trader has on marketplaces in a timely manner.

5. **Experimental setup**

We use JCAT 0.17¹ to run CAT games to simulate competing marketplaces with traders moving between them, so as to examine the effectiveness of the marketplace

¹<http://jcat.sourceforge.net/>.

selection strategies we described in the previous section. In particular, we are interested in comparing these strategies and finding out how they be configured to perform most effectively in terms of, for instance, the profits of traders and the social welfare.

5.1. Marketplaces

In our experiments, each marketplace runs the *continuous double auction* or CDA (Friedman, 1993) and adopts one of the following charging policies that were introduced and explored by (Niu et al., 2010, 2007):²

- *Fixed charging* (GF): charges a fixed portion of profit of trader in each transaction, where the profit is defined as the difference between the price offered by the trader and the transaction price.
- *Bait-and-switch charging* (GB): lowers its charge on the profit of a trader to attract more traders to trade in its marketplace whenever its market share is below a pre-fixed threshold and then gradually increases the charge to make more profit until the market share falls below the threshold again.
- *Charge-cutting charging* (GC): sets its charge to a fraction of the lowest charge imposed by other marketplaces based on information from the previous day if the marketplace is not charging the least among all the marketplaces. Unlike GB, GC never tries to increase charges.
- *Learn-or-lure-fast charging* (GL): sets its charge based on its belief about the exploration of traders in choosing marketplaces. If the traders are spread among marketplaces more or less evenly, just as what happens at the beginning of a game, the marketplace adjusts its charge towards 0 to lure the traders that are believed to be still exploring. On the contrary, if the traders appear to have concentrated in one or two marketplaces, it is assumed that traders have learned enough information and would be reluctant to leave the marketplace they found the best, so the marketplace, instead of luring traders, adjusts its charge towards that of the most profitable marketplace. The adjustment is made gradually, following the Widrow-Hoff rule that traders use to learn their expected utilities.

These charging policies are used in our experiments to provide market mechanisms that differ from each other.

Broadly, we perform two distinct sets of experiments, with five marketplaces in each experiment. In the first set of experiments, all marketplaces are *non-adaptive*. That is, all traders use GF and charge at different levels (0%, 20%, 40%, 60%, and 80% respectively on the profit of traders). In the second set of experiments, all marketplaces are *adaptive*,

²In our experiments, we only consider the charges on the profit of traders, although JCAT allows a marketplace to charge traders in various ways.

using one of the three adaptive charging policies (two with GB,³ two with GC,⁴ and one with GL,⁵ imposing charges initially in the same way as the non-adaptive ones do). These two compositions allow us to explore how marketplace selection strategies perform in different scenarios.

5.2. Traders

We run 100 trading agents in each CAT game, which are evenly split between sellers and buyers. All the traders use the *zero-intelligence constrained* (ZI-C) strategy in making offers. In this strategy, traders randomly choose a shout price, with the exception that they will not make a loss; e.g. a buyer will not place a bid that is above its private value. The naïvety of ZI-C allows us to discount the effects of the shout strategy on the results.

The traders homogeneously use one of the marketplace selection strategies that we described earlier to choose a marketplace — that is, in each experimental run of the simulation, all traders are using the same marketplace selection strategy with the same parameters. These are the independent variables in our experiment. To investigate the optimal values of the parameters used by these marketplace selection strategies, we pick a set of discrete values in the range of each parameter and, for each strategy, we run two experiments with each combination of its parameter values, one with non-adaptive marketplaces and the other with adaptive marketplaces. Table 1 lists all the parameters in the marketplace selection algorithms, their ranges, and the intervals we choose to determine the discrete values for these parameters. Taking into account all combinations, we run a total of 1,642 experiments.⁶

Each experiment runs 30 CAT games and each game lasts 400 days. That is, for each of the 1642 cases, 30 independent games are run. The results to be presented in the next section are averaged over the total 12,000 days. Traders are each allowed to trade one unit of goods each day and their private values are drawn from the uniform distribution between 50 and 100.

5.3. Measures

In these experiments, we record several measures, of which two are of particular interest.

First, we measure the *mean trader profit*, which is the mean daily profit over all simulations of all traders for all days. This provides us with a way to measure the

³GB sets the *threshold* on market share to 0.3 and the *cut ratio* — the speed of decreasing the charge — to 0.9 in our experiments.

⁴GC scales down from the lowest charge imposed on the previous day by a factor of 0.8 in our experiments.

⁵GL uses the so-called *single day exploring monitor* in JCAT with a threshold of 0.6 to decide whether the population of traders are still exploring or not.

⁶ $1642 = 2 \times (20 + 20 + 400 + 380 + 1)$, see the last column in Table 1.

Table 1: Values of parameters in the experiments.

Strategy	Parameter	Min	Max	Interval	Cases
ϵ -greedy	ϵ	0.00	0.95	0.05	20
ϵ -first	ϵ	0.00	0.95	0.05	20
ϵ -decreasing*	ϵ	0.00	0.95	0.05	400
	α	0.00	0.95	0.05	
softmax [†]	τ	0.01	1000.0	exponential	380
	α	0.05	0.95	0.05	
random	–	–	–	–	1

* In our experiments, the minimum value of ϵ is fixed at 0.

† In our experiments, the minimum value of τ is fixed at 0.01.

effectiveness of a marketplace selection strategy over a long period of trading. The daily profit for a trader i is:

$$pr_i = \begin{cases} |v_i - p_i| - f_i & (\text{where } p_i > 0) \\ -f_i & (\text{where } p_i = 0) \end{cases} \quad (1)$$

in which v_i is the private valuation of trader i , p_i is the price of the trade made by trader i , and f_i are the fees paid by trader i . In the case that a trader does not make a successful trade that day, they lose the fees charged by the marketplace, which would be 0 in the case that the only fees are a cut of the profit, as is the case in our experiments.

The mean daily profit of all traders on a single day is:

$$P = \frac{\sum_i^N pr_i}{N} \quad (2)$$

in which N is the number of traders.

Second, we measure the *global allocative efficiency*, which measures how close the entire market is to trading at the equilibrium price, where the *equilibrium price* is defined as the price at which demand equals supply when all traders offer to buy or sell at their private value, assuming that all traders in the market can trade with each other. The global allocative efficiency is calculated using:

$$E = \frac{\sum_j \sum_i |v_i^j - p_i^j|}{\sum_j \sum_i |v_i^j - p_0|} \quad (3)$$

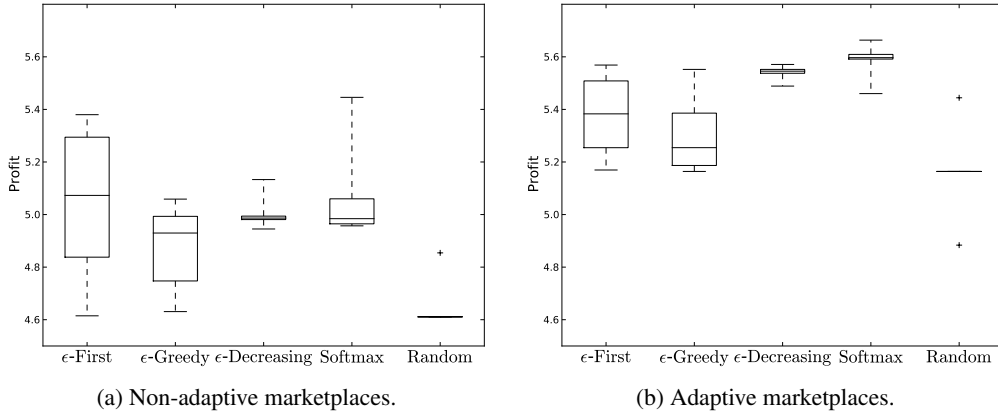


Figure 1: Five-number summaries of mean daily profit over all simulations for each strategy.

in which p_0 is the equilibrium price of the market, v_i^j is the private value of trader i in marketplace j , and p_i^j is the price paid by trader i in marketplace j .

These two measures, trader profit and global efficiency, allow us to assess how marketplace selection strategies affect the individual traders as well as the global market. Trader profit is of interest to individual traders, while global allocative efficiency is of interest to people designing a competitive market. Global allocative efficiency is also of interest to designers creating a market-based solution to resource allocation in a closed system.

6. Results

In this section, we present the results of the experiments outlined in Section 5.

6.1. Trader profit

Figure 1 shows the five-number summaries of mean daily trader profit for each strategy; that is, each data point is the mean of an explored parameter value. For the random strategy, which has no parameters, outlier markers ('+') are used to represent the 95% confidence interval for the random samples. These summaries provide an overview of the “stability” of a strategy. While the optimal parameter settings demonstrate the “best” choice for the 30 iterations we have run, the stability may allow us to generalise the performance over more iterations, and different experimental parameters.

From these figures, one can see that softmax returns the highest trader profit for both adaptive and non-adaptive marketplaces. Figure 1 demonstrates that ϵ -first is the least consistent, with large intervals in the five-summary spread. ϵ -decreasing is a highly consistent strategy in which the parameters have minimal impact. The high number of outliers is due to the small interval between the first and third quartile values.

The most notable result from this data is that the random strategy achieves a lower mean payoff than almost all combinations of strategies and parameters, making it evident that an intelligent marketplace-selection strategy is beneficial for traders.

The figures in Appendix A provide a more detailed view of the parameter space for each of the strategies assessed, except for the random strategy, which has no parameters.

Figure A.3 plots the ϵ parameters against mean daily trader profit for the ϵ -first and ϵ -greedy algorithms. The 95% confidence intervals are plotted as error bars. Figures A.4 and A.5 plot the profit results for ϵ -decreasing and softmax, which both contain two parameters. Figure A.5 shows the data in a 3D plot, while Figure A.4 shows the same data, except from one dimension only – that is, the average profit for α over all values of ϵ in Figure A.4a, and the average profit for ϵ over all values of α in Figure A.4b. The error bars show the *average* of all of the 95% confidence intervals of these parameters.

For ϵ -first and ϵ -greedy, in both types of market, the mean profit peaks with the parameters in the range 0.05–0.2, and then steadily decreases as ϵ approaches 1.0, which is consistent with expectations that only a small amount of exploration is required Sutton and Barto (1998). A large value for ϵ increases the probability that a random action is chosen, and our results indicate that a random strategy provides a poor payoff for this problem. Additionally, there is a large reduction in profit for ϵ -first at $\epsilon = 0.55$, which is consistent across all iterations of our experiments. We have no explanation for this.

In non-adaptive markets, ϵ -first achieved a high trader profit when ϵ was in the range 0.05–0.2, which is to be expected, as it learns the optimal choice in the first $\epsilon \times A$ rounds, and then exploits this for the remainder of the game. ϵ -greedy achieves less profit because it continues to explore even after the optimal marketplace is clear.

As expected, ϵ -first and ϵ -greedy return lower trader profit in the adaptive market, relative to ϵ -decreasing and softmax. This is due to the different ways of handling the exploration vs. exploitation problem. ϵ -first returns a lower profit because it does not explore after $\epsilon \times A$ rounds, and therefore cannot adapt to changes in the market. ϵ -greedy returns a lower profit because it continues to explore late in the game, even when the optimal choice is clear. In contrast, both ϵ -decreasing and softmax will adapt by continuing to explore late in the game, but will exploit the optimal choice far more often than ϵ -greedy.

There is also a marked difference in profit between adaptive and non-adaptive marketplaces, even for the random marketplace selection traders. This is because non-adaptive marketplaces do not change their fees, while adaptive marketplaces do. As a result, adaptive marketplaces will drive down their fees to attract traders, providing more profit to the traders.

Figures A.4 and A.5 show that, for the ϵ -decreasing, higher values of both ϵ and τ return higher profit for non-adaptive marketplaces, and the small spike at these points in Figure A.5a demonstrates that the combination of high values for these is the most profitable for traders. However, the effect of the parameters is small. For adaptive marketplaces, neither parameter has a major effect on profit.

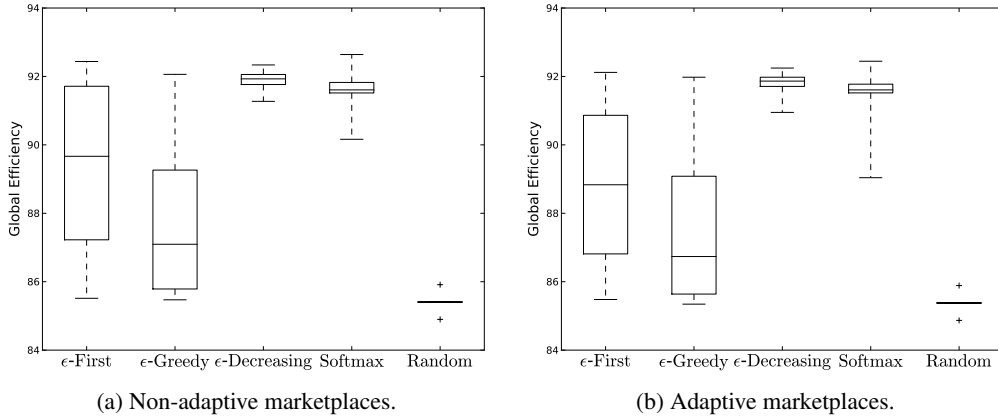


Figure 2: Five-number summaries of mean global efficiency over all simulations for each strategy.

For softmax, one can see that high values for α and τ return higher trader profit, more so in non-adaptive than adaptive markets. The combination of higher values implies a close-to-random exploration strategy for the initial days, then slowly moving towards a strategy of exploitation for the remainder.

To maximise trader profit overall, the results suggest that in both markets, a softmax strategy should be chosen. In a non-adaptive market, the profit peaked at $\alpha = 0.95$ and $\tau = 10$. In an adaptive market, the profit peaked at $\alpha = 0.8$ and $\tau = 10$.

6.2. Global market efficiency

Figure 2 shows the five-number summaries of mean global efficiency for each strategy. The data points are the same as for Figure 1, except the measurement is global market allocative efficiency instead of trader profit.

From this figure, one can see that ϵ -decreasing achieves the highest median efficiency for either adaptive or non-adaptive markets, but that softmax achieves the highest single average. In fact, there is little difference in scores between the adaptive and non-adaptive markets for all strategies. This is contrast to the trader profit measures, which show a higher profit in adaptive markets. The same effect is not seen in the global efficiency measure because the calculation outlined in Section 5 does not consider fees in their calculation — only prices. Thus, competition between marketplaces does not lead to higher global efficiency in these experiments. In this figure, we see the same outliers as in Figure 1 for softmax and ϵ -decreasing.

The figures in Appendix Appendix B provide a more detailed view of the parameter space. Figure B.6 plots the ϵ parameters against mean efficiency for the ϵ -first and ϵ -greedy algorithms. From this figure, one can see that in both types of market, the efficiency peaks at low values of ϵ in both algorithms, and then demonstrate a clear

trend downwards, which is consistent with the profit results in Section 6.1.

Figures B.8 and B.7 plot the efficiencies for the ϵ -decreasing and softmax algorithms, with the former as a 3D plot, analogous to the plots in Section 6.1. These plots provide little indication of the most efficient parameters for ϵ -decreasing, but for softmax, higher values of τ and α are preferred for softmax.

To maximise efficiency, the results suggest that a softmax strategy should be used in both adaptive and non-adaptive markets. In a non-adaptive market, efficiency peaked for the softmax algorithm at $\alpha = 0.85$ and $\tau = 5$. The softmax algorithm achieves the top 10 highest scores in non-adaptive markets. In adaptive markets, efficiency peaked at $\alpha = 0.6$ and $\tau = 500$.

6.3. Threats to Validity

We identify two major threats to validity in our experimental approach.

The first major threat is that we have used homogeneous marketplace selection strategies. That is, in each experiment run, all traders used the same strategy with the same parameters. In an open market, this is unlikely to be the case, and the interplay between these different strategies may result in some algorithms performing better or worse under certain parameter settings than in the homogeneous case. However, for system designers attempting to efficiently allocate resources in a closed system, it is possible that a homogeneous strategy will be used.

The second major threat is the small parameter space explored in the experiments. While the parameter spaces of the strategies were explored systematically and thoroughly, other possible market setup parameters were held constant, such as the number of marketplaces, the traders' shouting strategies, the number of traders, and the number of days in a CAT game. It is clear that changing such parameters can impact individual traders' profit and the market in general. However, the results are in-line with our expectations that adaptive strategies increase trader profit and global efficiency, and that the exploration phase should be small compared to the exploitation phase. Furthermore, the strategies are implemented as general N-armed bandit solutions, so we are confident that the results generalise for other market setups.

6.4. Discussion

From the results, it is clear that the softmax strategy is preferred from an individual trader perspective and a global perspective. We attribute the success of the softmax strategy to the Boltzman distribution used to select the next action. This distribution increases the amount of exploration in a balanced market, but has two properties:

1. the best marketplace is still chosen more than any other marketplace; and
2. when choosing a "non-optimal" marketplace, the exploration is more likely to choose marketplaces in which the agent has been successful in the past, rather than a random selection.

One interesting discussion point is the relative ranking of strategies between trader profit and global efficiency. These two measures are both related to the price of traders, with the difference being that trader profit has fees subtracted.

Figures 1a and 2a show an interesting difference between trader profit and efficiency in non-adaptive markets. For trader profit, ϵ -decreasing is the third-ranked strategy for both the best value and the median value. However, for global efficiency, ϵ -decreasing ranks the highest for median, and is the most stable over its parameter space.

These results indicate that in comparison with other strategies, for the ϵ -decreasing strategy, the high global allocative efficiency is not to the benefit to the individual traders, whose profits are low in comparison to other strategies. This implies that the marketplaces themselves are making higher profits (relative to other strategies) via fees. Therefore, the ϵ -decreasing strategy is good for resource allocation in non-adaptive markets, but not for the individual traders in these markets. This result is not found in adaptive markets due to the competition driving prices down.

7. Conclusion

In this paper, we assess several N-armed bandit algorithms for solving the problem of marketplace selection in double-auction markets with competing marketplaces. The algorithms assessed are: 1) the ϵ -first algorithm; 2) the ϵ -greedy algorithm; 3) the ϵ -decreasing algorithm; and 4) the *softmax* algorithm (Sutton and Barto, 1998). In addition, we assess a *random* choice algorithm as a baseline.

Overall, the results indicate that the softmax strategy is the most successful at maximising trader profit and global allocative efficiency in both adaptive and non-adaptive markets. The ϵ -decreasing strategy performs similarly well in adaptive markets, and is more stable than softmax over its parameters space. Furthermore, ϵ -decreasing achieves high and stable global allocative efficiency in non-adaptive markets, but this is not to the benefit to the individual traders, whose profits are low in comparison to other strategies. This implies that the efficiency leads to higher profits for marketplaces (relative to other strategies) via fees.

The results also indicate, for both trader profit and efficiency, that the exploration phases of the learning algorithms should be relatively small compared to the exploitation; around 5%–15% of all actions, which is consistent with previous work on reinforcement learning (Sutton and Barto, 1998). This holds whether the exploration phase is at the start of the trading game (as in ϵ -first) or throughout the game.

By comparing the results of the strategies to random marketplace selection, the results demonstrate that an intelligent marketplace selection strategy is better for both trader profitability and market efficiency. This extends the work of Cai et al. (2008), who demonstrated that allowing traders to migrate improved the efficiency of a market. Our results demonstrate that choosing the right marketplace selection strategy and parameters can lead to higher market efficiency, as well as increasing trader profit.

In future work, we will address the issue on heterogeneous strategies; that is, cases in which traders are employing different marketplace selection algorithms. In addition, we aim to build on the work of Sohn et al. (2009) by investigating marketplace-specific trading strategies, which integrate the marketplace selection strategy with the trading strategy itself.

References

- Brafman, R. I., Tennenholtz, M., 2003. R-MAX – a general polynomial time algorithm for near-optimal reinforcement learning. *The Journal of Machine Learning Research* 3, 213–231.
- Cai, K., Niu, J., Parsons, S., 2008. On the economic effects of competition between double auction markets. In: *Proceedings of the 10th Workshop on Agent-Mediated Electronic Commerce (AMEC 2008)*.
- Friedman, D., 1993. The double auction institution: A survey. In: Friedman, D., Rust, J. (Eds.), *The Double Auction Market: Institutions, Theories and Evidence*. Santa Fe Institute Studies in the Sciences of Complexity. Westview Press, Perseus Books Group, Cambridge, MA, Ch. 1, pp. 3–25.
- Ganchev, K., Nevmyvaka, Y., Kearns, M., Vaughan, J. W., 2010. Censored exploration and the dark pool problem. *Communications of the ACM* 53 (5), 99–107.
- Gode, D. K., Sunder, S., 1993. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy* 101 (1), 119–137.
- Greenwald, A. R., Kephart, J. O., 1999. Shopbots and pricebots. In: *Agent Mediated Electronic Commerce (IJCAI Workshop)*. pp. 1–23.
- Ladley, D., Bullock, S., 2005. Who to listen to: Exploiting information quality in a ZIP-agent market. In: *Proceedings of IJCAI-05 Workshop on Trading Agent Design and Analysis (TADA-05)*.
- Niu, J., Cai, K., Parsons, S., Gerding, E., McBurney, P., Moyaux, T., Phelps, S., Shield, D., 2008. JCAT: A platform for the TAC market design competition. In: *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems: Demo papers. IFAAMAS*, pp. 1649–1650.
- Niu, J., Cai, K., Parsons, S., McBurney, P., Gerding, E., 2010. What the 2007 TAC Market Design Game tells us about effective auction mechanisms. *Journal of Autonomous Agents and Multiagent Systems* 21 (2), 172–203.
- Niu, J., Cai, K., Parsons, S., Sklar, E., 2007. Some preliminary results on competition between markets for automated traders. In: *Proceedings of AAAI-07 Workshop on Trading Agent Design and Analysis (TADA-07)*.
- Rochet, J. C., Tirole, J., 2003. Platform competition in two-sided markets. *Journal of the European Economic Association* 1 (4), 990–1029.
- Shi, B., Gerding, E., Vytelingum, P., Jennings, N., 2010. A game-theoretic analysis of market selection strategies for competing double auction marketplaces. In: *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS*, pp. 857–864.
- Sohn, J., Lee, S., Mullen, T., 2009. Impact of misalignment of trading agent strategy under a multiple market. In: *Proceedings of the First Conference on Auctions, Market Mechanisms and Their Applications*. Springer, pp. 40–54.
- Sutton, R. S., Barto, A. G., 1998. *Reinforcement learning: An introduction*. MIT press, Cambridge, MA.
- Whittle, P., 1988. Restless bandits: Activity allocation in a changing world. *Journal of Applied Probability* 25, 287–298.
- Widrow, B., Hoff, M. E., 1960. Adaptive switching circuits. *IRE WESCON Convention Record* 4, 96–104.

Appendix A. Trader Profit Results

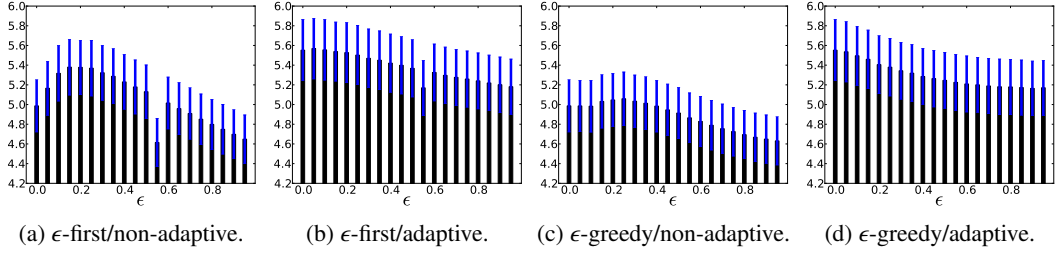


Figure A.3: Mean trader profit for the ϵ -first and ϵ -greedy algorithms.

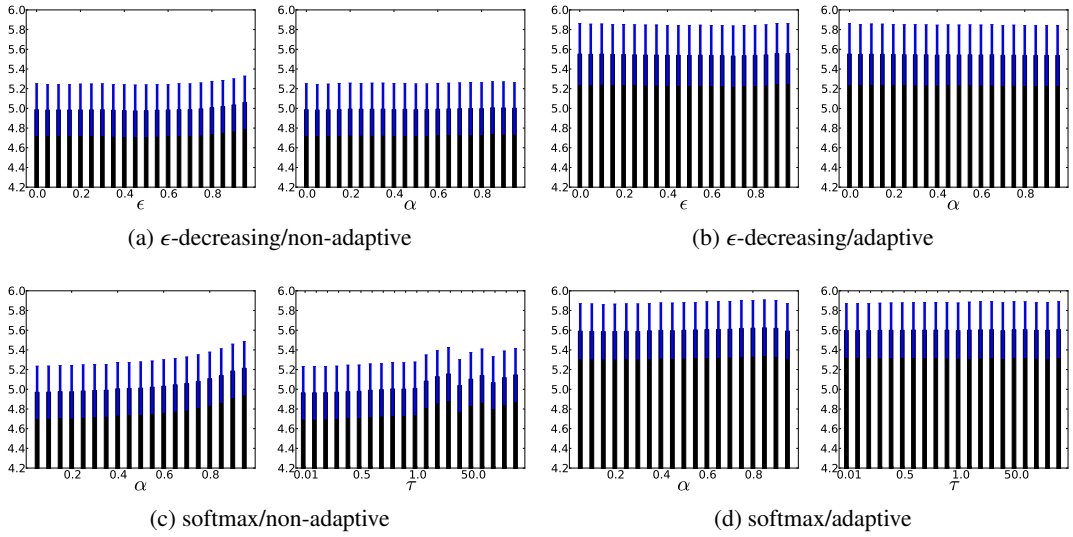
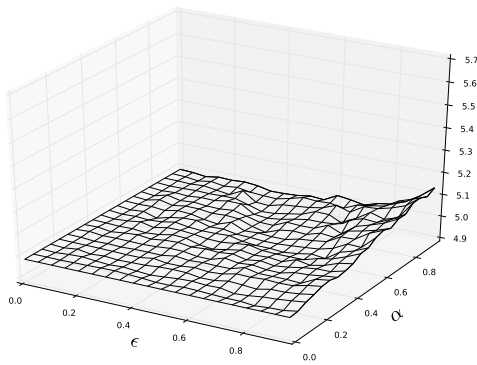
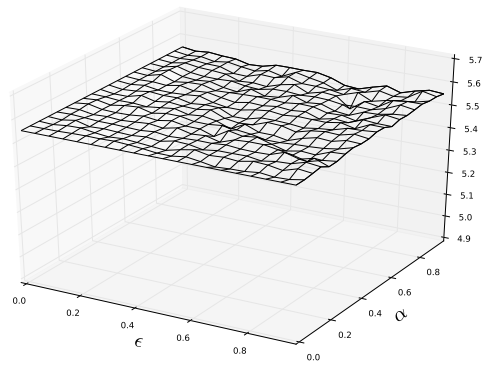


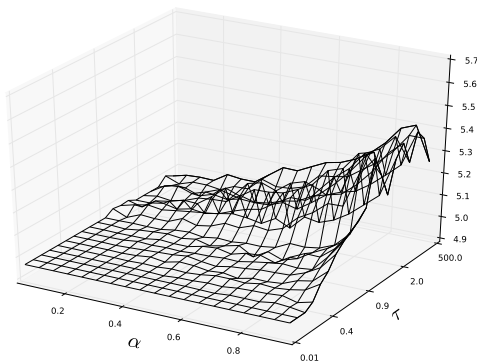
Figure A.4: Plots of mean trader profit of the ϵ , τ , and α parameters for the ϵ -decreasing and softmax algorithms.



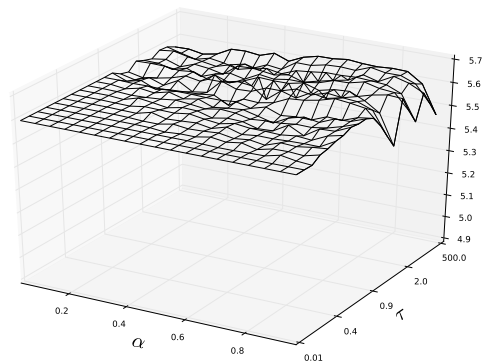
(a) ϵ -decreasing/non-adaptive.



(b) ϵ -decreasing/adaptive.



(c) softmax/non-adaptive.



(d) softmax/adaptive.

Figure A.5: 3D plots of mean trader profit for the ϵ -decreasing and softmax algorithms.

Appendix B. Global Efficiency Results

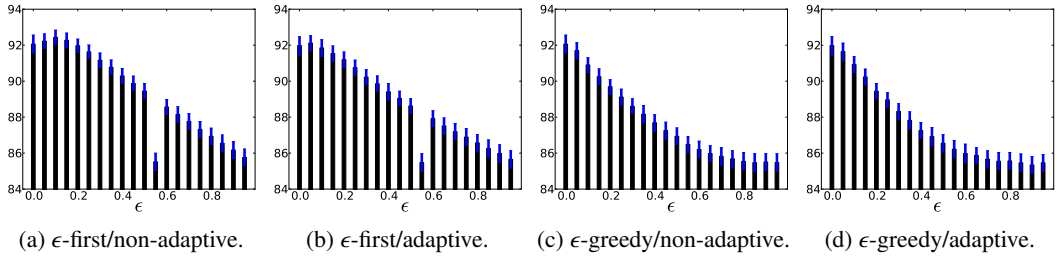


Figure B.6: Mean global efficiency for the ϵ -first and ϵ -greedy algorithms.

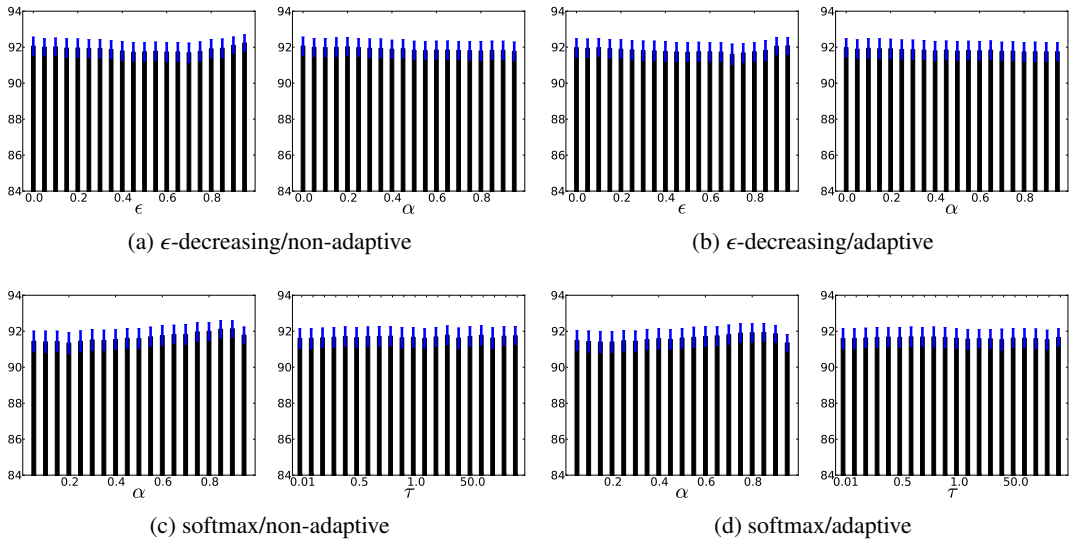
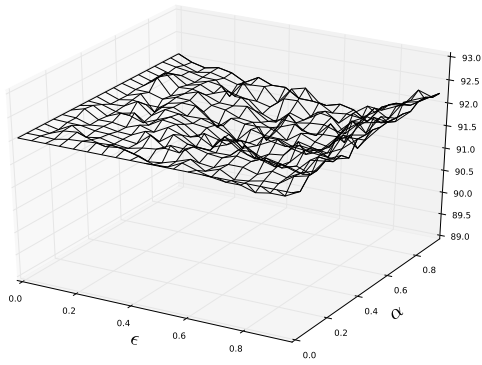
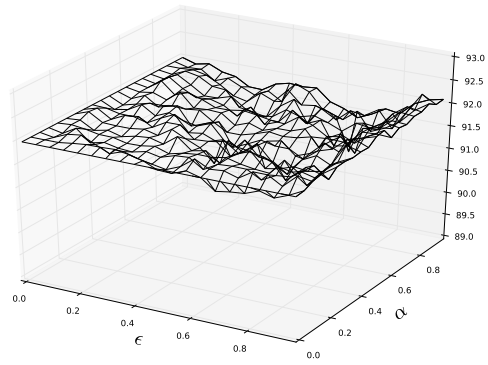


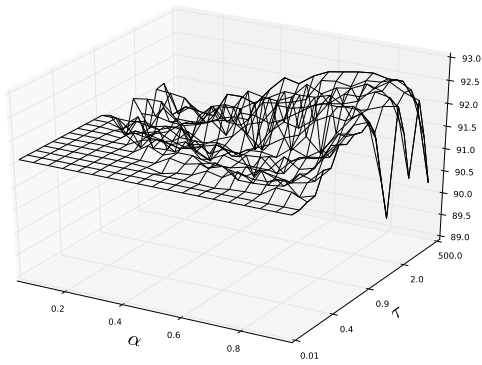
Figure B.7: Plots of mean global efficiency of the ϵ , τ , and α parameters for the ϵ -decreasing and softmax algorithms.



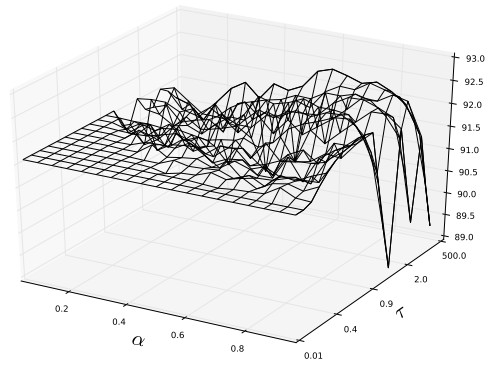
(a) ϵ -decreasing/non-adaptive.



(b) ϵ -decreasing/adaptive.



(c) softmax/non-adaptive.



(d) softmax/adaptive.

Figure B.8: 3D plots of mean global efficiency for the ϵ -decreasing and softmax algorithms.