# THE COMPARISON OF PRICING METHODS IN THE CARBON AUCTION MARKET *VIA* MULTI-AGENT Q-LEARNING

Akram Esmaeili Avval, Farzad Dehghanian* and Mohammadali Pirayesh

**Abstract.** In this paper, the uniform price and discriminative price methods are compared in the carbon auction market using multi-agent Q-learning. The government and different firms are considered as agents. The government as auctioneer allocates initial permits in the carbon auction market, and the firms as bidders compete with each other to obtain a larger share of the auction. The carbon trading market, penalty, reserve price, and bidding volume limitation are considered. The simulation analysis demonstrates that bidders have different behavior in two pricing methods under different amounts of carbon permits. In the uniform price, the value of bidding volume, firms' profit, and the trading volume for low permits and the value of the government revenue, clearing price, the trading price, and auction efficiency for high permits are greater than ones in the discriminative price method. Bidding prices have a higher dispersion in the uniform price than the discriminative price method for different amounts of carbon permits.

## 1. Introduction

Reducing the amount of carbon emissions has been a major concern in many countries. To curb carbon emissions, some policies are adopted by governments [8]. Cap and trade scheme is admitted as one of the most efficient schemes by many researchers like Du *et al.* [15] and Cao *et al.* [6]. Under a cap and trade scheme or emission trading scheme (ETS), the firms are allocated some carbon permits from the government, and they can trade carbon permits in the carbon trading market.

One of the most important issues in ETS is which allocation patterns should be implemented [4,10]. According to the existing literature, the carbon allowance allocation methods are free allocation (grandfathering approach and benchmarking approach), auction, and a combination of the above approaches. Due to the disadvantages of free allocation (distortion of the carbon price, windfall profits, etc.), auctions should be considered a preferable way to allocate carbon permits [9]. For instance, in the third phase of EU-ETS, 48% of carbon permits are distributed by auction, and it is predicted that the proportion for auction will be 90% by 2021–2030 [19]. Given the superiority of the auction-based methods and their increasing applications, we focus on auction-based methods in this paper.

Industrial Engineering Department, Faculty of Engineering, Ferdowsi University of Mashhad, PO Box 91775-1111, Mashhad, Iran.
*Corresponding author: f.dehghanian@um.ac.ir

There are several types of auctions that are mainly divided into two groups: static (sealed) and dynamic (clock) [28]. In a static auction, two market pricing methods, *i.e.*, uniform pricing and discriminatory pricing, have been developed. In the uniform price method, each winner pays the market clearing price (the price at which the aggregate demand curve intersects the supply curve), while in the discriminatory price method, they pay their bid prices [12, 13].

Selecting the right pricing method for a carbon market is still a hot debate. For example, Cong and Wei [11] indicate when carbon allowances are nearly low, the discriminative price method is more efficient than the uniform price, while little participants benefit more in the other method. Cong and Wei [12] show when the number of participants is relatively high, and there is no communication between them, the English auction clock is more effective than the uniform and discriminatory pricing. But when the number of bidders is relatively low and there is communication between them, discriminatory price auction is better than two other methods and prevents collusion. Tang *et al.* [28] demonstrate the uniform price method has a smaller effect on economic damage and emission reduction compared to the discriminatory price method.

Many studies, like Santos *et al.* [7], Hattori and Takahashi [16], Sugiyarto [27], Akbari-Dibavar *et al.* [1], and Matthäus [20], have been conducted to compare these two pricing methods in different areas, but their results are not consistent. However, there are few studies that have compared these methods in the carbon auction market. Generally, in this area, various methods have been used to analyze the behavior of participants. For instance, Jiang *et al.* [19] investigate how market power can influence the auction price using mathematical models and some operations like derivative. They consider two allocations patterns, mixed allocation, and single auction. They examine the effect of these two patterns on compliance costs and welfare. Cong and Wei [12] utilize experimental method to compare some carbon auction methods. Dormady [14] provides some experiments to simultaneously investigate a carbon and energy market under real-world market characteristics.

Due to the lack of auction data and the nature of its gameplay, individual behavior in this system is complex [12]. Equations cannot model interactions in these systems [5], and experiments are expensive. One of the powerful tools in studying complex systems is the multi-agent-based model [31]. For instance, Cong and Wei [11] use this tool to compare some carbon auction methods. They consider the government as the auctioneer and two types of plants as the bidders. Tang *et al.* [28] design a carbon allowance auction market using the multi-agent-based model. They consider two agents: the government as the regulator of emission trading scheme and different firms in all parts of China. Yu *et al.* [21] propose a multi-agent-based model to simulate an emission trading scheme consisting of some firms in China.

Our work is an extension of the search conducted by Cong and Wei [11]. Their study was limited to the carbon auction market, while the carbon trading market and carbon trading price can affect carbon auction and *vice versa*. They used the Roth–Erev reinforcement learning algorithm to adjust the bidding strategies. Their algorithm presented a bidding strategy based on the result obtained from the past, while studies show that if the agents can anticipate the long-term outcomes of their current bidding strategy rather than optimizing their immediate rewards, their profitability will improve [29]. As well as, in their study, each agent could generate a bid price close to zero because of not using reserve price. Therefore, the permits might have been sold below their value. Reserve price is the minimum price that the government expects to be paid for a permit unit in the carbon auction. It can guarantee that the permits are not purchased below their real value [9].

This paper is going to compare the uniform pricing method and discriminative pricing method using the multi-agent-based model in which each adaptive agent illustrates a firm that takes part in carbon auction in a cap-and-trade scheme and determines its bid based on Q-learning. Watkins proposed the Q-learning algorithm to solve the Markovian Decision Problems with incomplete information. This algorithm has some features that make it appropriate for repeated games against unknown opponents. First, it does not need a model from the environment. Second, it can be utilized on-line to discover the optimal strategy using the experience gained from interacting directly with the environment [29, 32]. Third, an agent can predict the long-term outcomes of its actions and the actions of other agents, and therefore, it can be able to correctly model the other agents and achieve the optimal bidding strategy [29]. This algorithm has received increasing attention in the electricity auction market and has become a major tool in solving this problem [23, 24, 32, 33]. As mentioned, the carbon

permit auction market is a complex economic system due to a lack of data. In this system, each agent's behavior is strongly dependent on other agents' behavior and market conditions. Besides, each of the agents faces a lack of knowledge about their other competitors. Under such circumstances, building a model for the economic system is a complex problem, and using free-model algorithms such as the Q-learning algorithm can be very appropriate.

The next section formulates the presented multi-agent-based model. Section 3 describes the Q-Learning algorithm and proposes agents' bidding strategies according to it. Section 4 presents the experimental results. Section 5 concludes the paper and offers direction for future research.

## 2. Problem definition

In this section, we present our multi-agent-based model to compare two pricing methods, namely the uniform pricing method and discriminative pricing method in a carbon auction market. We consider two kinds of agents, government, and firms. The government, as an auctioneer, allocates the initial carbon permits to the bidder and determines the actual price (the price that each firm should pay). Also, it regulates the carbon trading price to prevent market power and manipulating price. The firms, as bidders, simultaneously submit demand schedules (their prices and the quantities that they are willing to buy at those prices) to the auctioneer. The government determines the clearing price, the gained carbon permit from each firm, and the actual carbon price by forming the aggregate demand and supply curve. The government adds demand schedules to build the aggregate demand curve. The price in which the aggregate demand curve and supply curve intersect each other demonstrates the clearing price. The demands that are above the clearing price will be answered, the ones that are at the clearing price will be rationed, and the ones that are below the clearing price will be rejected. By comparing required carbon permits during production with carbon permits gained in the auction, each firm determines its supply or demand in the carbon trading market. Given the amount of supply and demand in this market, the government determines the carbon trading price. If required carbon permits exceed the carbon permits gained from two markets, the firm will be fined.

Table 1 defines the main parameters used throughout this paper.

### 2.1. Agents' decisions

As mentioned in the previous section, first, the government as auctioneer determines the initial carbon permit $TotalPermit$. Suppose the minimum carbon permit required for all firms to support their productions is $Totalemission$. To decrease the total emissions, the government implements the reduction policy and controls the initial carbon permit; therefore, it provides a percentage of $Totalemission$, i.e., $\phi$. Consequently, $TotalPermit$ is calculated based on equation (2.1):

$$TotalPermit = Totalemission * \phi. \tag{2.1}$$

The firms are the bidders in our model. Each firm $i$ needs $e_i$ permits to produce one unit of product. Suppose production cost and selling price per unit product for firm $i$ are $c_i$ and $p_i$, respectively. Therefore, the value of a permit to firm $i$ can be calculated as equation (2.2):

$$v_i = \frac{p_i - c_i}{e_i}. \tag{2.2}$$

In the equation above, $v_i$ is the private value (the maximum price that each firm tends to pay for a permit unit). In the carbon auction market, each firm gives a bid price $bp_i$ that is between reserve price and $v_i$ and a bid volume $bv_i$ (the required carbon permit to support its production). The government ranks firms in descending order in terms of their bid prices to form the aggregate demand curve. The equilibrium price is equal to the bid price generated by firm $k$ that satisfies the following inequalities:

$$\sum_{i=1}^{k} bv_i \leq TotalPermit \tag{2.3}$$

TABLE 1. The parameters definition.

| Notation | Explanation |
|---|---|
| $TotalPermit$ | The initial carbon permit |
| $Totalemission$ | The minimum carbon permit required for all firms to support their productions |
| $\phi$ | The percentage of carbon permit supply |
| $e_i$ | The required carbon permit for producing one unit product by firm $i$ |
| $c_i$ | Production cost per unit product for firm $i$ |
| $p_i$ | Selling price per unit product for firm $i$ |
| $v_i$ | Private value for firm $i$ |
| $bp_i$ | The bid price for firm $i$ |
| $bv_i$ | Bid volume for firm $i$ |
| $ep$ | The equilibrium price in the carbon auction market |
| $gv_i$ | Carbon permit gained in auction for firm $i$ |
| $cp$ | Market clearing price |
| $ap_i$ | The price paid for each carbon permit unit by firm $i$ |
| $rv_i$ | The total carbon permit required by firm $i$ |
| $nv_i$ | Surplus or insufficient carbon permits of firm $i$ |
| $PermitSupply$ | Total permit supply in carbon trading market |
| $PermitDemand$ | Total permit demand in carbon trading market |
| $nv'_i$ | The actual carbon permit traded by firm $i$ in carbon trading market |
| $ctp$ | Carbon trading price |
| $Srevenue$ | Sales revenue for firm $i$ |
| $Ccost_i$ | The costs of carbon permits in the auction and carbon trading market for firm $i$ |
| $Pun_i$ | The penalty paid by firm $i$ for exceeding its emissions from the gained permit |
| $Q_i$ | The product sales amount by firm $i$ |
| $Temission_i$ | Total emissions generated by firm $i$ |
| $Gpermit_i$ | Total permits obtained by firm $i$ in carbon auction and carbon trading market |
| *Algorithm parameters* | |
| $S = \{s_1, s_2, \ldots, s_n\}$ | The set of the environment's state |
| $A = \{a_1, a_2, \ldots, a_m\}$ | The set of the agent's actions |
| $Q(s, a)$ | $Q$ value for each permissible pair $(s, a)$ |
| $\pi^*(s)$ | The optimal policy |
| $P_{ss'}(a)$ | Possibility of changing the environment' state from $s$ to $s'$ by selecting action $a$ |
| $R^a_{ss'}$ | The immediate reward of the agent for taking action $a$ in state $s$ and changing environment' state to $s'$ |
| $T = \{0, 1, \ldots\}$ | The steps that the algorithm repeat |
| $s_t$ | The environment's state in step $t$ |
| $a_t$ | The agent' action in step $t$ |
| $r_t$ | The immediate reward obtained by the agent in step $t$ |
| $\gamma$ | Discounted factor |
| $\alpha$ | Learning rate |
| $\epsilon$ | A small probability |

$$\sum_{i=1}^{k+1} bv_i > TotalPermit. \tag{2.4}$$

Therefore, the equilibrium price $ep$ is equal to $bp_k$, *i.e.*, $ep = bp_k$. The carbon permits gained of firm $i$ ($gv_i$) in the carbon auction market for two pricing method is calculated according to the following equation:

$$gv_i = \begin{cases} bv_i & bp_i > bp_k \\ \frac{bv_i\left(TotalPermit - \sum_{i=1}^{j-1} bv_i\right)}{\sum_{i=j}^{k} bv_i} bp_j = \ldots bp_i = \ldots = bp_k, & j \le i \le k \\ 0 & bp_i < bp_k. \end{cases} \tag{2.5}$$

In the uniform price method, every winner pays the market clearing price $cp$ that is equal to the equilibrium price $ep$. Therefore, the actual carbon price $ap_i$ (the price that firm $i$ should pay) in the uniform pricing method is calculated according to equation (2.6):

$$ap_i = \begin{cases} cp & i \leq k \\ \infty & i > k. \end{cases} \tag{2.6}$$

In the discriminative price method, the winners pay their bidding price. So, the actual carbon price $ap_i$ in the discriminative pricing rule is calculated according to equation (2.7):

$$ap_i = \begin{cases} bp_i & i \leq k \\ \infty & i > k. \end{cases} \tag{2.7}$$

The clearing price $cp$ in the discriminative price method is calculated based on equation (2.8):

$$cp = \sum_{i=1}^{k} bp_i bv_i \bigg/ \sum_{i=1}^{k} bv_i . \tag{2.8}$$

After the firms obtain the initial carbon permits in the carbon auction market, they can trade their carbon permits in the carbon trading market. If carbon permits obtained for firm $i$, $gv_i$, in the carbon auction is greater than its required carbon permits $rv_i$, the firm has surplus carbon permits, i.e., $nv_i = rv_i - gv_i < 0$. Therefore, the firm $i$ can sell them in the carbon trading market. Similarly, if $gv_i$ is less than $rv_i$, the firm has insufficient permits, i.e., $nv_i = rv_i - gv_i \geq 0$, so they should purchase carbon permits required in the carbon trading market. Total permit supply and total permit demand in carbon trading marked shown by $PermitSupply$ and $PermitDemand$, respectively, are calculated according to the following equations

$$PermitSupply = \sum_{nv_i < 0} -nv_i \tag{2.9}$$

$$PermitDemand = \sum_{nv_i > 0} nv_i \tag{2.10}$$

when the total permit supply is greater than total permit demand, i.e., $PermitSupply \geq PermitDemand$, the carbon trading price decreases and the actual permit traded by firm $i$, $nv_i'$ is obtained by equation (2.11). Otherwise, the carbon trading price increases and $nv_i'$ is calculated according to equation (2.12).

$$\begin{aligned} nv_i' \\ PermitSupply \geq PermitDemand \end{aligned} = \begin{cases} 0 & nv_i = 0 \\ \frac{PermitDemand * nv_i}{PermitSupply} & nv_i < 0 \\ nv_i & nv_i > 0 \end{cases} \tag{2.11}$$

$$\begin{aligned} nv_i' \\ PermitSupply < PermitDemand \end{aligned} = \begin{cases} 0 & nv_i = 0 \\ \frac{PermitSupply * nv_i}{PermitDemand} & nv_i > 0 \\ nv_i & nv_i < 0. \end{cases} \tag{2.12}$$

## 2.2. Determining carbon price in the carbon trading market

The carbon trading price has a significant effect on the carbon auction. If the carbon trading price is high, the bidders might attempt to buy more permits in the auction to sell them in the carbon trading market at a higher price. If the carbon trading price is low, the competition between the bidders reduces, and consequently, the clearing price decreases. To maintain stability and avoid manipulation in the carbon market, the government considers the carbon trading price $ctp$ around the market clearing price within a given range $\beta \sim U(0, 0.3)$.

$$ctp = \begin{cases} (1 + \beta) * cp & PermitSupply \leq PermitDemand \\ (1 - \beta) * cp & PermitSupply > PermitDemand. \end{cases} \tag{2.13}$$

## 3. Agents' bidding strategy

In this section, we present a profitable bidding strategy based on the Q-learning algorithm. First, we propose the Q-learning algorithm and then describe the proposed bidding strategy.

### 3.1. Q-learning algorithm

Suppose an agent interacts with its environment at discrete time steps, $t = 0, 1, 2, \ldots$, and $S = \{s_1, s_2, \ldots, s_n\}$ is a finite set of states that the environment can adopt, and $A = \{a_1, a_2, \ldots, a_m\}$ is a finite set of actions that the agent can select. At each time step $t$, the agent observes the present state of the environment $s_t = s \in S$ and then takes an action $a_t = a \in A$. Consequently the agent achieves an immediate reward $r_{t+1}$ and the environment alters its states to new states $s_{t+1} = s' \in S$ according to the transition probability $p_{ss'}(a)$.

In the Q-Learning algorithm, there is a lookup table containing $Q$ value for each permissible pair $(s, a)$. First, the table is initialized either randomly or according to the agent's knowledge. The purpose of the agent is to find the optimal policy $\pi^*(s) \in A$ to maximize $Q$ value of each state over the long run by using the Bellman optimality:

$$Q^*(s, a) = \sum_{s'} P_{ss'}(a) \left[ R_{ss'}^a + \gamma \max_{a'} Q^*(s'a') \right] \tag{3.1}$$

$$\pi^*(s) = \arg\max_a \left( Q^*(s, a) \right). \tag{3.2}$$

In equation (3.1), $\gamma \, (0 \leq \gamma \leq 1)$ is the discounted factor and determines how important the future rewards are for the agents. $R_{ss'}^a$ is an immediate reward that the agent receives because of taking action $a$ in state $s$ and changing the environment' state from $s$ to $s'$. Without knowing $p_{ss'}(a)$, the Q-learning algorithm is capable of finding the optimal policy for each state by online estimation of $Q(s, a)$ in a recursive method and by utilizing data: $s_t$, $a_t$, $s_{t+1}$ and $r_{t+1}$. The updating equation is as follows:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha \Delta Q_t(s_t, a_t) \tag{3.3}$$

$$\Delta Q_t(s_t, a_t) = \left\{ r_{t+1} + \gamma \max_{a'} [Q_t(s_{t+1}, a')] \right\} - Q_t(s_t, a_t) \tag{3.4}$$

where $\alpha \, (0 < \alpha < 1)$ is the learning rate and represents the degree of new data impact on the update of estimated $Q$ values. In other word, $\alpha$ represent how much the agents consider the recent information to explore possibilities. Based on the theorem of the QL convergence, the $Q$-value converges to the optimal value, if each $(s, a)$ is visited infinitely and $\alpha$ decreases appropriately. In a single agent, the environment is stationary and Markovian, but our problem is a multi-agent case that an agent's optimal policy depends on other agents' strategy. Therefore, each agent provides a non-stationary and non-Markovian environment for the other agents, and these conditions do not guarantee convergence to the optimal policy. Some progress has been made in this area [17, 18, 26], but they are for some special cases and seem inappropriate for practical problems. Despite these, among all the reinforcement learning algorithms, Q-learning is applied in many studies mainly because of its simplicity.

### 3.2. Q-learning based bidding strategy

In auctioning carbon permits, firms try to determine their bid prices and bid volumes to maximize their profits. Based on the Q-learning algorithm, every firm learns from its bidding experience during runs to bid profitably. To determine the profitable bidding strategy for firms using the Q-learning algorithm, the states of the environment, actions, and rewards must first be specified.

The market clearing price is considered as the state of the environment. That changes between the reserve price value and the maximum price that the firms can afford to pay per unit of permit. To prevent "the curse of dimensionality" the state space is equally discretized into 16 states.

Deciding on bidding price and volume is the action of each agent. The bidding price for firm $i$ is between the reserve price and $v_i$, and bidding volume is between the minimum permits that the firm requires for supporting

its productions and the maximum permits that the firm is allowed to bid. The price space and volume space are equally discretized into 6 prices and 16 volumes, respectively. Therefore, the action space is $6 * 16$.

The reward of firm $i$ participating in carbon permit auction is its benefit function that is calculated as follows:

$$\pi_i = Srevenue_i - Ccost_i - Pun_i \tag{3.5}$$

where $Srevenue_i = (p_i - c_i) * Q_i$ is sales revenue (where $c_i, p_i$ and $Q_i$ present production cost and selling price per unit product and the sales amount of firm $i$, respectively). $Ccost_i$ displays the costs of carbon permits in the auction and carbon trading market and is calculated according to equation (3.6):

$$Ccost_i = ap_i * gv_i + ctp * nv'_i. \tag{3.6}$$

$Pun_i$ presents the penalty paid by firm $i$. If total emissions generated by firm $i$ ($Temission_i$) exceed the total permits obtained from the carbon auction and carbon trading market ($Gpermit_i = gv_i + nv'_i$), it should pay the penalty for its non-compliance emissions.

### 3.3. Algorithm implementation

According to Figure 1, the steps of firms' learning and bidding are given as follows:

(1) *Initialization*: first, the input parameters of the algorithm are initialized. Small random numbers or 0 are assigned to all state-action combinations for each firm.
Suppose, Maxiter is the maximum iteration intended to run the algorithm. It is a termination condition; therefore, the following steps are repeated until the termination condition is reached.

(2) *State identification*: in each iteration, the agents utilize the market clearing price on the previous step as the current state. In the first iteration, the reserve price (the lowest possible value) is considered the environment's state.

(3) *Action selection*: after identifying the environment's state, the agents select their actions (bidding decisions) according to $Q$-values saved in $Q$-value lookup tables for each state-action pair. To choose action, the agents utilize the $\epsilon$-greedy method to balance exploitation and exploration. According to this method, the agents select the action with maximum $Q$-value in the state $s$ with high probability $1 - \epsilon$ and a random action from all admissible actions with a small probability $\epsilon$.

(4) *Q-value update*: after declaring the market clearing price and carbon permit allocated to each firm by the government, the firms calculate their rewards according to equation (3.5) and update the $Q$-values according to the current rewards and next state, which is the market clearing price of the current iteration using equation (3.3) and (3.4).

## 4. EXPERIMENTAL ANALYSIS

In this section, based on the multi-agent $Q$-learning, the uniform price auction and discriminative auction are compared. Therefore, the algorithm parameters, data, and the results of the simulation are presented in the next two sub-sections.

### 4.1. Parameter setting

To compare two auction formats, it is supposed that five adaptive agents compete in the auction carbon market and trade their permits in the carbon trading market and explore their bidding strategy. The parameters of agents are randomly generated from the interval shown in Table 2.

As can be from Table 2, the lower and upper bounds of each parameter are close to each other; therefore, there is a perfectly competitive market, and no agent possesses the market power. The Q-learning parameters $\alpha$, $\gamma$ and $\epsilon$ is considered 0.9, 0.1 and 0.5, respectively. These values have an essential role in exploring and convergence of the Q-learning algorithm, Therefore, the related values are selected based on the previous studies conducted in this context, like Sadr *et al.* [25] and Poursalimi Jaghargh and Mashhadi [22], in a way that a balance between exploration and exploitation is obtained. To ensure that permits are not sold below their value, we set the reserve price at 0.5.
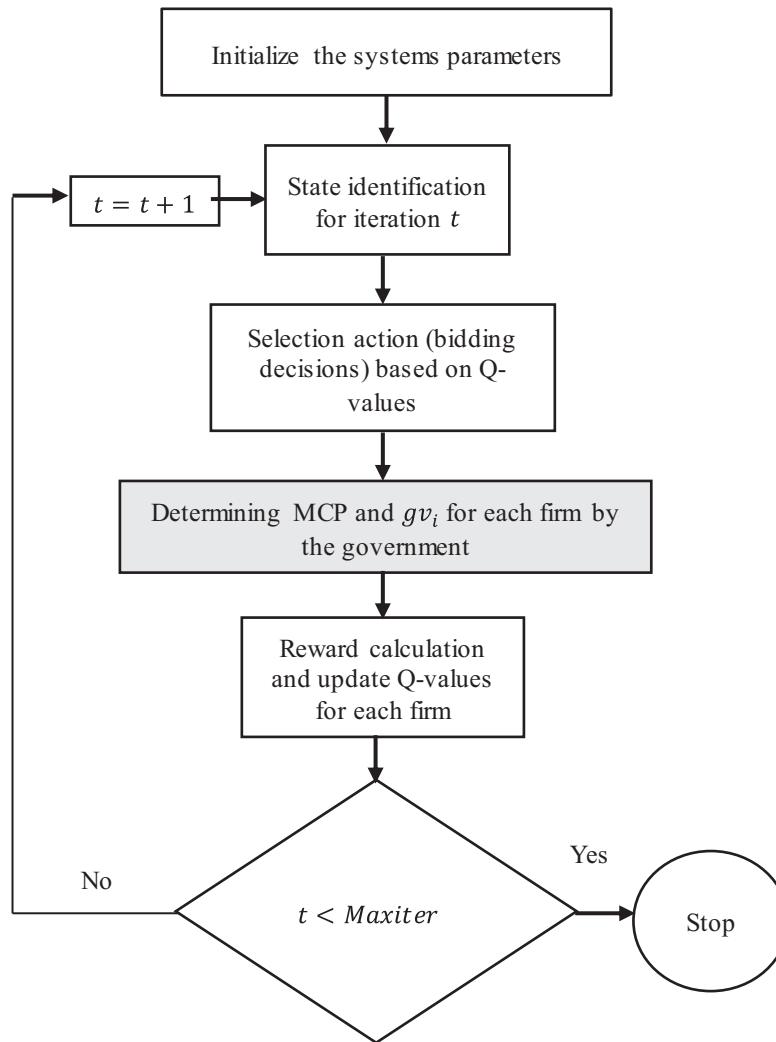
FIGURE 1. The scheme of QL-based bidding strategy.

TABLE 2. Parameters of the agents.

| The parameter | Value | The parameter | Value |
|---|---|---|---|
| $p_i$ | [7,9] | $e_i$ | [1.5, 2] |
| $c_i$ | [12, 14] | $Q_i$ | [100,150] |

## 4.2. Experiment results

In this section, we compare the uniform price and discriminative price method in terms of bidding price, bidding volume, firms' profit, the government revenue, the clearing price, the carbon trading price, the total amount of permits traded in the carbon trading market, the firm's benefit reduction in the carbon auction relative to the free method and emission reduction. Finally, we compared the performance of the Q-Learning

FIGURE 2. The equilibrium price and bid prices of all agents in the uniform pricing rule.

algorithm with the Roth–Erev algorithm that was used in Cong and Wei [11], in terms of firms' benefit. The algorithm has coded in C++ and complied with Microsoft Visual Studio 2012.

The $\phi$ value can influence the results; therefore, we compare these two price methods for different $\phi$ values. To investigate our results for low to high $\phi$ values, we change it from 80% to 100%. According to equation (2.1), when the value of $\phi$ is small, the carbon supply quantity is small and *vice versa*.

### 4.2.1. Impact of auction form on bidding price

Figures 2 and 3 give the equilibrium price and bid prices of all agents under the uniform and discriminative pricing rules, respectively. As we can see from the two figures, the agents bid prices are much higher than the equilibrium price in the uniform price auction, while, in the discriminative price auction, they bid prices as close to equilibrium price as possible. That is because, in the discriminative price, every winner pays its bid. Hence, the bidders tempt to predict the equilibrium price and bid close to it, but in the uniform price, since every winner pays the equilibrium price, therefore, forecasting the equilibrium is less important [13]. This fact makes the bid prices in the uniform auction have a higher dispersion [2, 3, 30]. For example, in our experiment, the average variances for bid prices in the uniform and discriminative price are 0.423 and 0.05 price unit, respectively. As shown in the two figures, the bid price of agent 2 is almost equal to the equilibrium price for all values of $\phi$.

All firms bid below their private values in both methods. It can be seen in Figure 4. Note with the increase in the value of $\phi$, the bid prices decrease. This is because $\phi$ affects supply quantity. When the supply quantity is low, the competition between firms increases; therefore, they bid higher prices and close to their private values. But, when the supply quantity increases, the competition decreases, so the firms bid as close to the reserve price as possible. As you can see, the bid prices are more reduced in the discriminative price method, and they are almost equal to 0.5 (reserve price) for $\phi = 90\%$ to 100% for all firms. As previously mentioned, the bid price of firm 2 is equivalent to equilibrium price for $\phi = 80\%$ to 98%, because this firm wins some of its bid; therefore, it is not economically feasible that it bids a higher price. From $\phi = 98\%$ to 100%, firm 2 increases its price in the uniform price because other firms decrease their bid prices, and it allows firm 2 to increase its bid price and win more than before.

### 4.2.2. Impact of auction form on bidding volume

As shown in Figure 5, the bid volumes fluctuate around a fixed value in the uniform price method. Because, in this method, the firms try to influence the equilibrium price and keep it down. While with the increase in $\phi$,
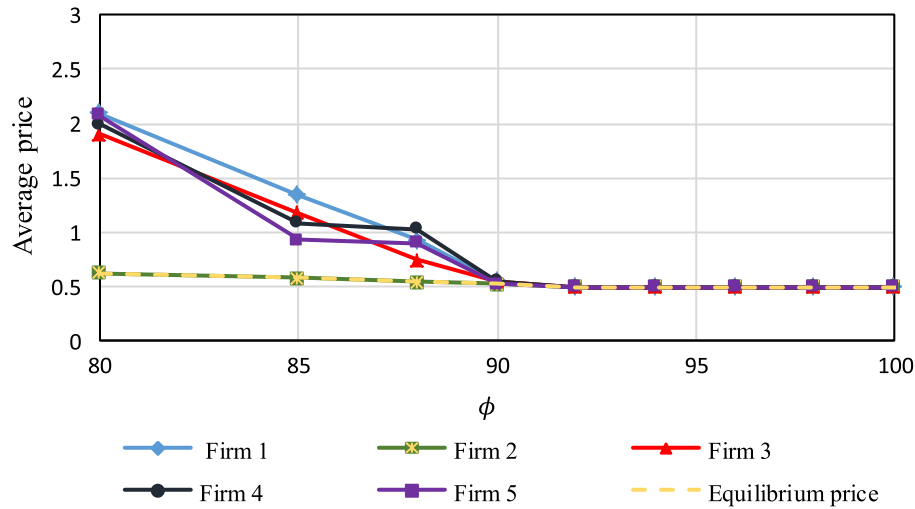
FIGURE 3. The equilibrium price and bid prices of all agents in the discriminative pricing rule.

bids volume increase in the discriminative pricing method. There are two reasons for this: first, in this method, firms pay their bids and, second, with the increase in $\phi$, according to Figure 4, the bid prices decrease so the firms can increase their bid volume to raise their benefits. Therefore, we can say for the high value of $\phi$, bid shading in the discriminative price is less than the uniform price method.

### 4.2.3. Impact of auction form on firms' profit

The benefit of firms is influenced by $\phi$ as shown in Figure 6. For $\phi$ less than 90%, the benefit of firms in the uniform pricing is more than the discriminative pricing method. This is because the firms pay the equilibrium price in the uniform auction. But for $\phi$ greater than 90%, the benefit of firms in the discriminative pricing is equal to or greater than the benefit of firms in the uniform pricing. Because, as shown in Figure 4, the bid prices are too low and almost equal to the equilibrium price for $\phi$ greater than 90%, therefore the benefit of firms increase. It can be concluded that when supply is scarce, the uniform price auction is more beneficial to bidders than discriminative price and *vice versa*.

### 4.2.4. Impact of auction form on the government revenue

Figure 7 shows, when the carbon permit is scarce the government's revenue in the discriminative price method is greater than that in the other method. Therefore when the carbon permit is large, the government should utilize the uniform price method, and when the carbon permit is scarce, the government should use the discriminative price method [11].

### 4.2.5. Impact of auction form on the clearing price

Figure 8 shows that the supply quantity can influence the clearing price. In the uniform price method, the clearing price is almost stable. But when the value of $\phi$ is less than 90%, the clearing price in the discriminative price is greater than that in the uniform price method, and for $\phi$ larger than 90%, it is less in the discriminative price method. Because with the increase in the value of $\phi$ in the discriminative price, the bid prices decrease and fall near the reserve price, so the clearing price reduces. In the uniform price method, the clearing price is equal to the equilibrium price, and as you can see in Figure 2, the equilibrium price is almost stable.
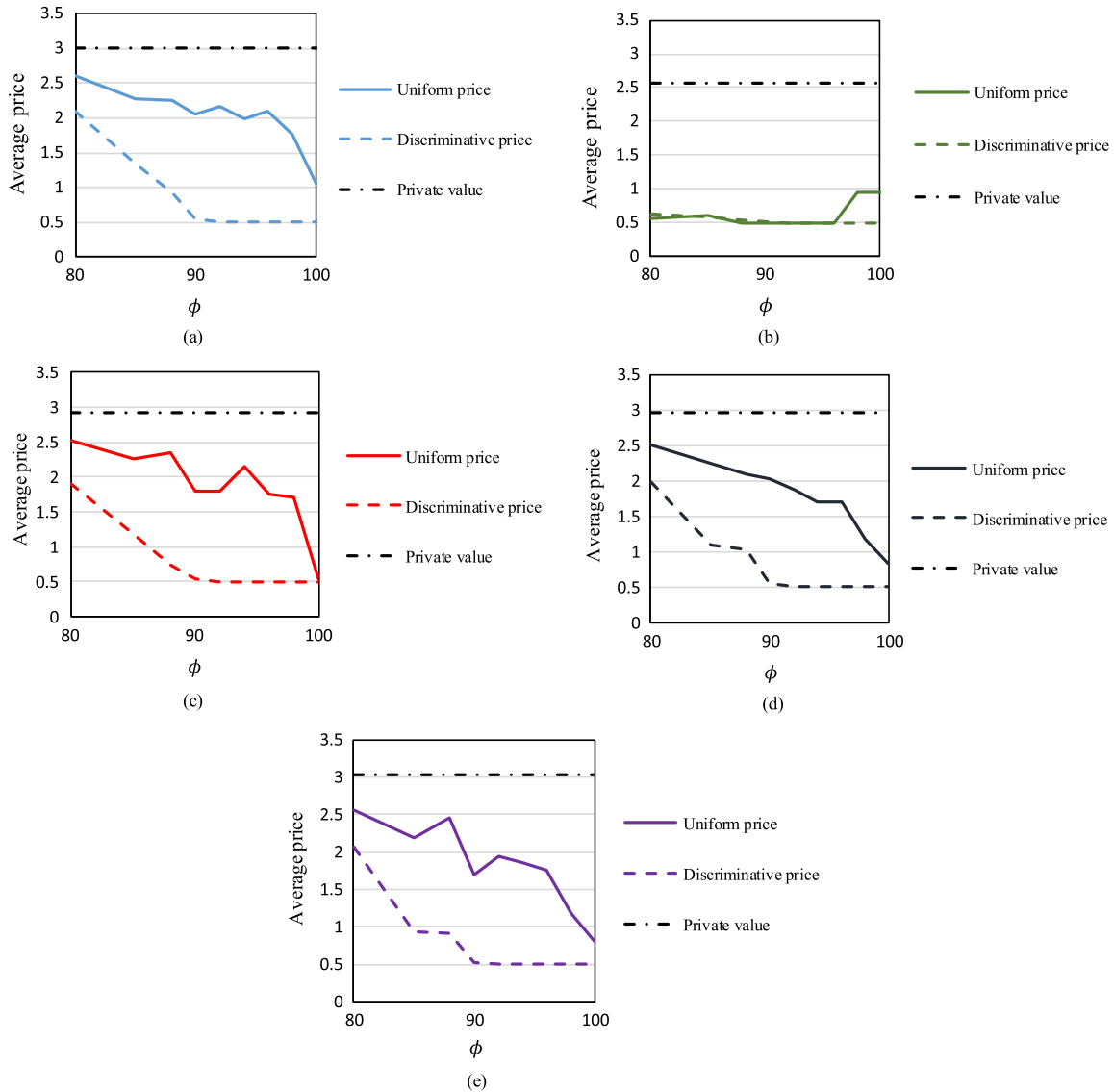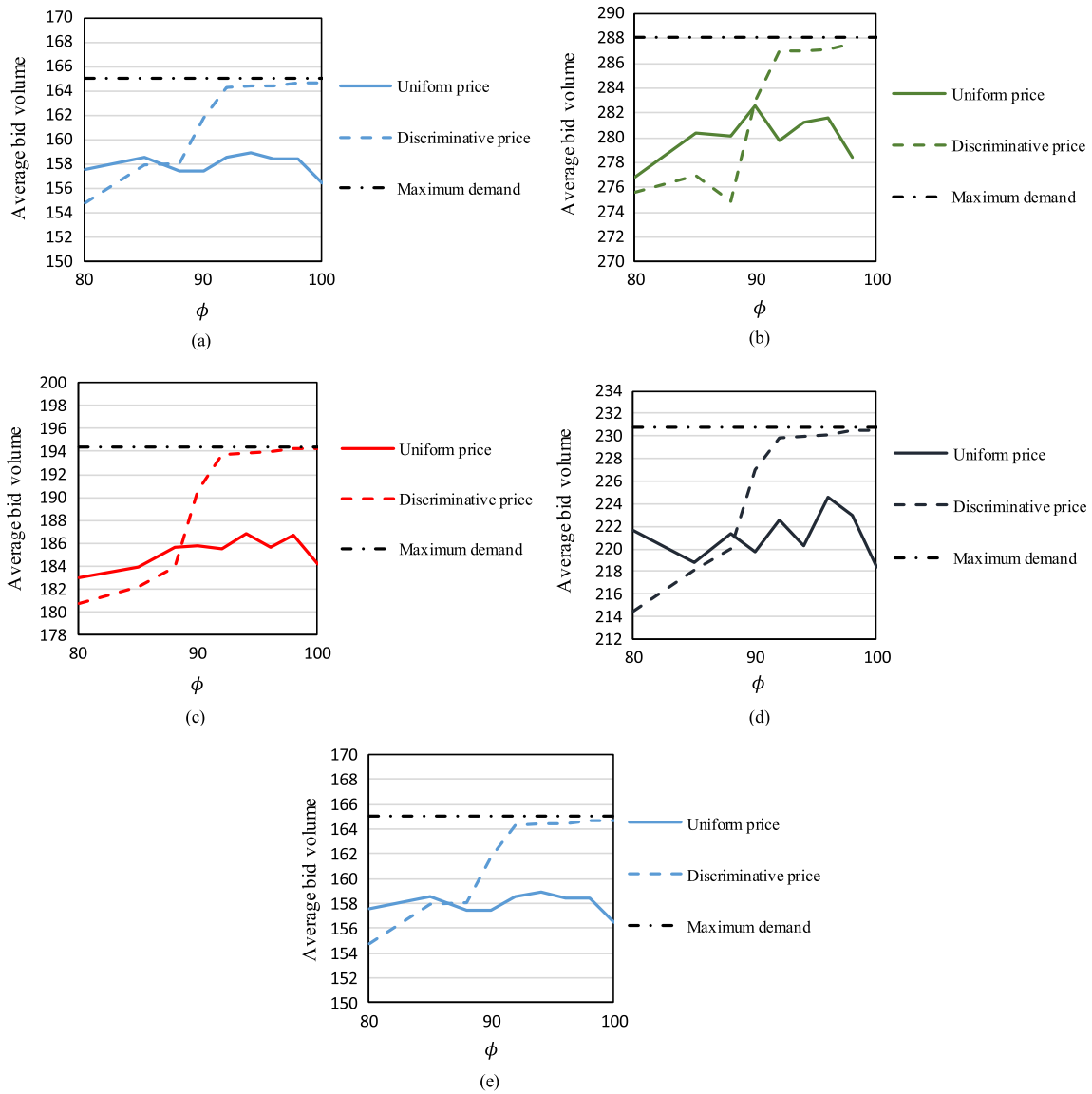
FIGURE 4. The comparison of bidding price in the two pricing method for different supply quantities: (a) Firm 1, (b) Firm 2, (c) Firm 3, (d) Firm 4, (c) Firm 5.

### 4.2.6. Impact of auction form on the carbon trading price

As mentioned before, the government determines the carbon trading price using the clearing price. Therefore, the relationship between the carbon trading price and supply quantity in the two pricing methods will be similar to the relationship between the clearing price and supply quantity. You can see it in Figure 9.

### 4.2.7. Impact of auction form on the carbon permit traded in the carbon trading market

The volume of carbon permits traded in the carbon trading market is different for the two pricing methods. As is evident from Figure 10, with the increase in supply quantity, the volume of trading increases in the discriminative price, while it is almost stable in the uniform price method. As before mentioned, the bidders
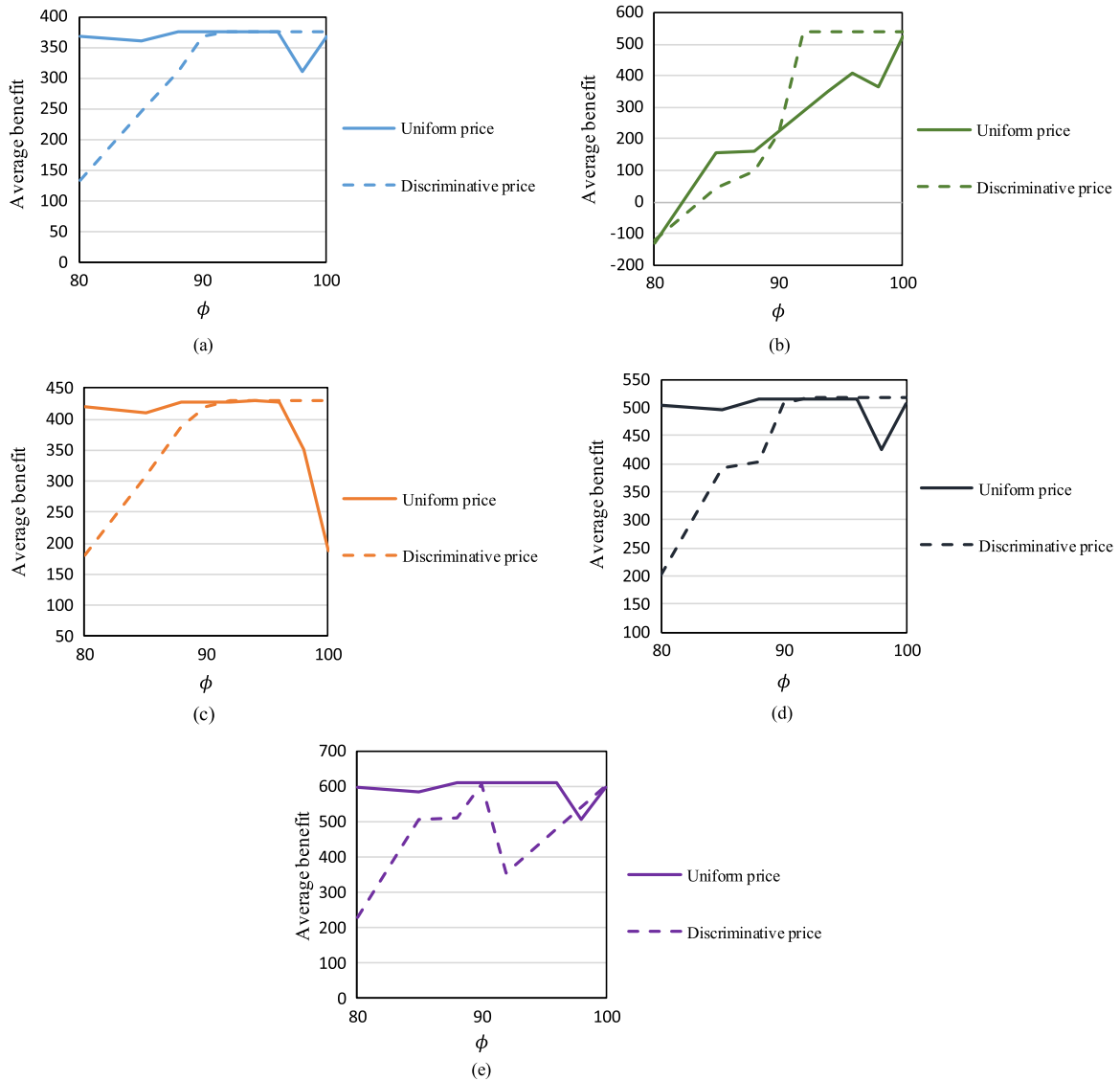
FIGURE 5. The comparison of bid volume in the two pricing methods for different supply quantities: (a) Firm 1, (b) Firm 2, (c) Firm 3, (d) Firm 4, (c) Firm 5.

in the uniform price keep their volume of bids as low as possible to influence the equilibrium price, but in the discriminative price, with the increase in supply quantity, the bidders increase their volume of bids to resell them at secondary price and raise their revenues.

### 4.2.8. Impact of auction form on the auction efficiency

In this section, we consider the auction efficiency as a performance measure. It can be calculated as follows [12]:

$$\mathrm{eff}_i = \frac{\mathrm{EOG}_i}{(\mathrm{EOG}_i + \mathrm{TBF}_i)}. \tag{4.1}$$

FIGURE 6. The comparison of firms' benefit in the two pricing methods for different supply quantities: (a) Firm 1, (b) Firm 2, (c) Firm 3, (d) Firm 4, (c) Firm 5.

In the above equation, $EOG_i$ is the actual earning of the government from auction and penalties. $TBF_i$ shows the total benefit of the firms. In two parameters, $i$ represents the type of auction. As shown in Figure 11, we definitely can not say which method is more efficient than the other. It depends on the supply quantity. When the supply quantity is low, the discriminative pricing method is more efficient than the uniform pricing method, but in case that the supply quantity is high, the uniform pricing method is more efficient than the other.

### 4.2.9. Impact of auction form on firms' benefit and emission reduction

In this section, our analysis focuses on the comparison of firms' benefit reduction in the auction compared to the situation that initial permit is allocated by free methods and emission reduction for the two pricing
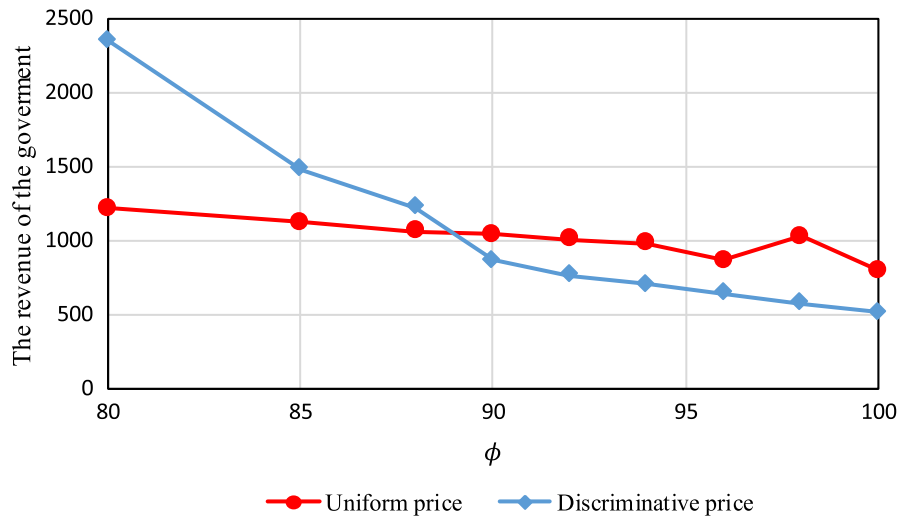
FIGURE 7. The revenue of the government for different supply quantities in the two pricing methods.
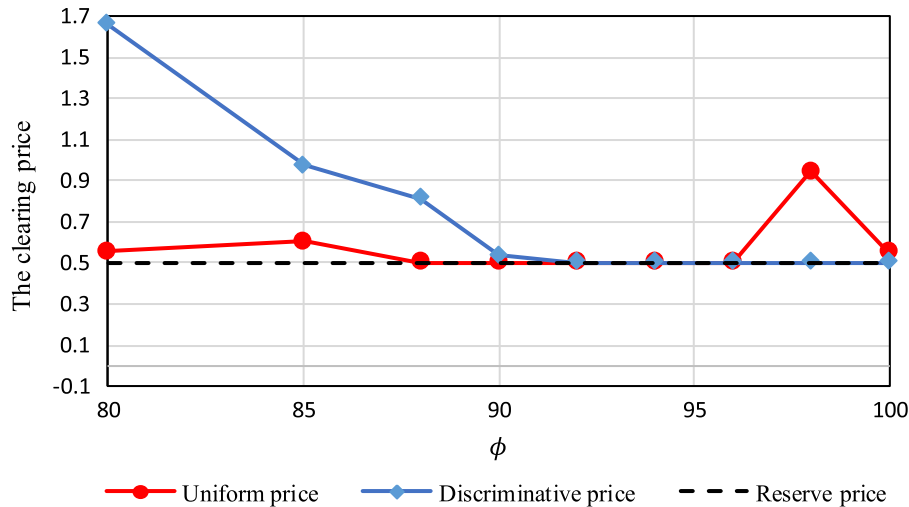


FIGURE 8. The clearing price in the two pricing methods.

methods. As shown in Figure 12, the values of firms' benefit reduction are positive for two methods; in other words, the carbon auction causes the firms' benefit to decrease. When $\phi$ is less than 90%, the amount of benefit reduction in the discriminative price is more significant than the uniform price and *vice versa*. This is because, in the uniform price, the firms earn more benefit than the other method for $\phi$ less than 90% (see Fig. 6).

As is evident from Figure 12, lowering the value of $\phi$ results in greater benefit reduction and emission reduction. The best environmental performance is obtained at $\phi = 80\%$, *i.e.*, when emission reduction is 20%. At this point, the most significant reduction in firms' benefit occurs. The smallest values of firms' benefit reduction in the uniform and discriminative price methods occur at $\phi = 96\%$ and $\phi = 100\%$, respectively. In other words, at these points, the best economic performances are obtained for two pricing methods. The results
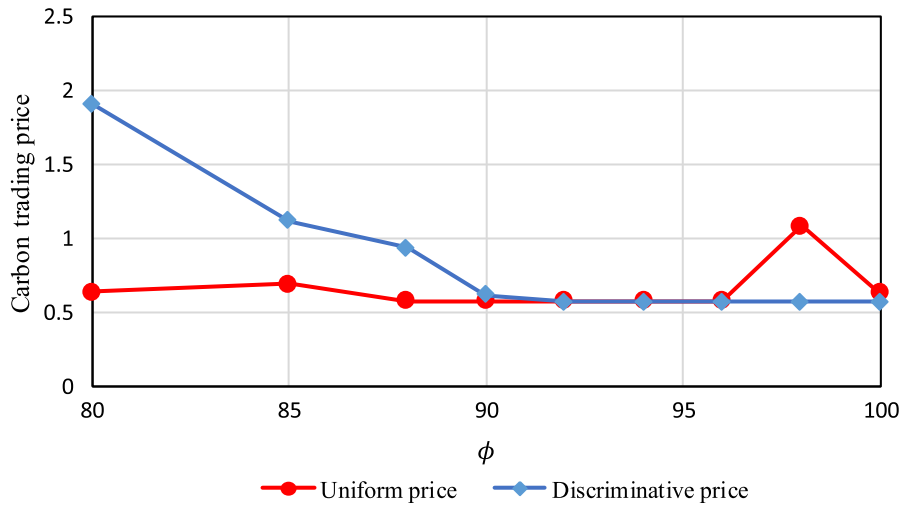
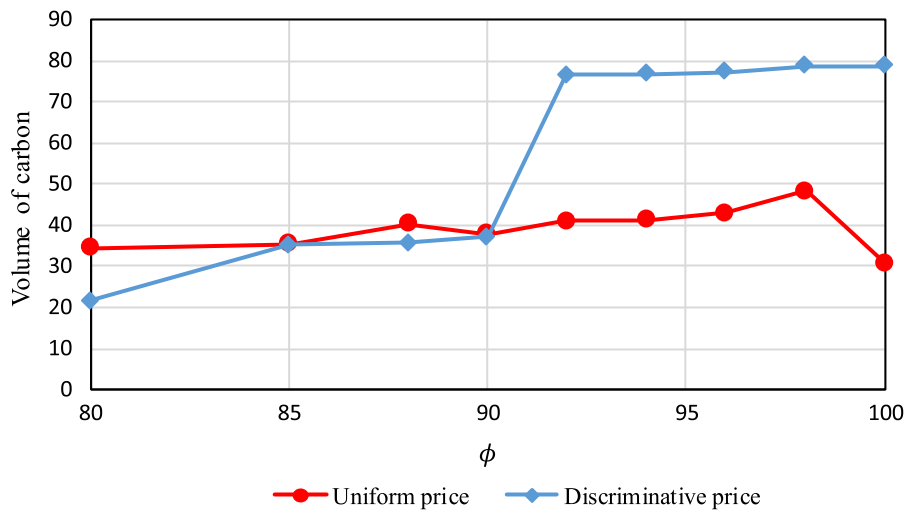FIGURE 9. The carbon trading price in the two pricing methods.



FIGURE 10. The volume of carbon traded in the carbon trading market in the two pricing methods.

of this section are appropriate for policymakers to determine the right value $\phi$ to balance between economic and environmental goals.

### 4.2.10. The comparison of Q-learning algorithm with Roth–Erev algorithm

As mentioned in Section 1, the Roth–Erev algorithm used by Cong and Wei [11] utilizes only past information and determines the bidding strategy. While if the agents can predict the long-term results of their current decisions, their profitability will increase. As said, the Q-learning algorithm can tackle this problem. In this section, we examined this claim and test the Roth–Erev algorithm and compare the results with the ones obtained from the Q-learning algorithm under the uniform and discriminative price for all values of $\phi$. As shown in Figure 13, the average firms' profit in the Q-learning algorithm outperforms the one in the Roth–Erev algorithm for all $\phi$ values and two pricing methods.
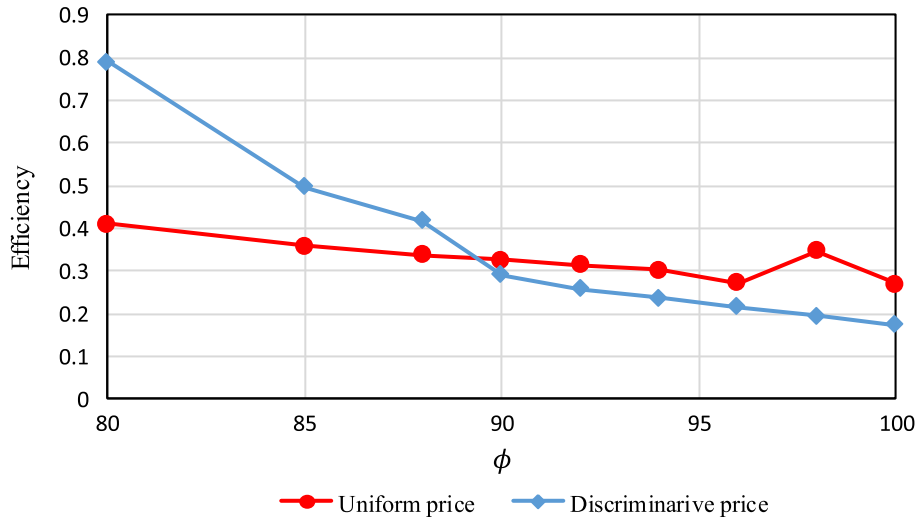
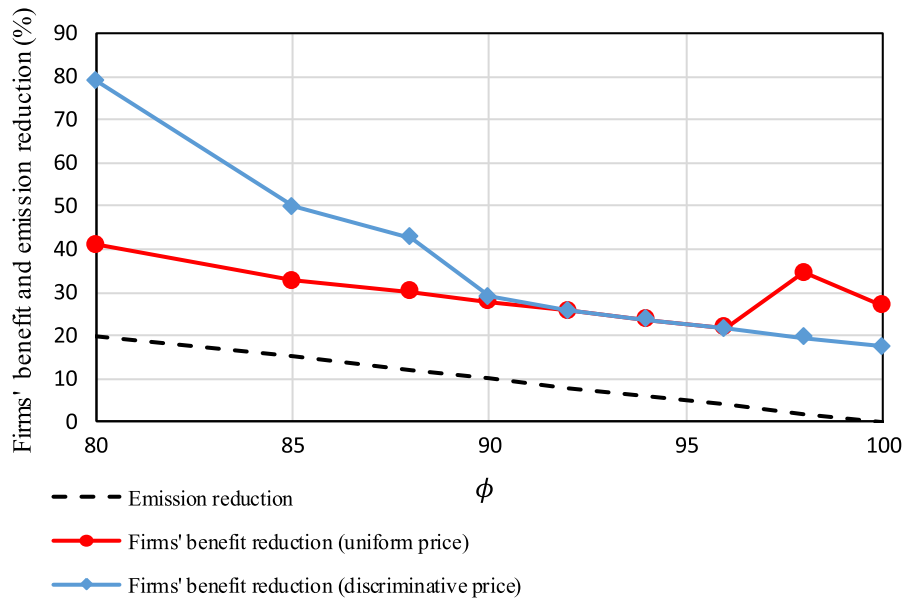FIGURE 11. Auction efficiency in the two pricing methods.



FIGURE 12. The firms' benefit reduction in the auction compared to the free method and emission reduction in the two pricing methods.

## 5. CONCLUSIONS

In this paper, we present a multi-agent-based model to compare the uniform price and the discriminative price methods in the carbon auction. Agents represent the firms participating in the auction, and they can develop their bids using the Q-learning algorithm. Our findings show that the agents' bidding behaviors are different in the two approaches. We cannot say which method is more advantageous, but the results are remarkably close to the theoretical predictions. The main of our conclusions are as follows:
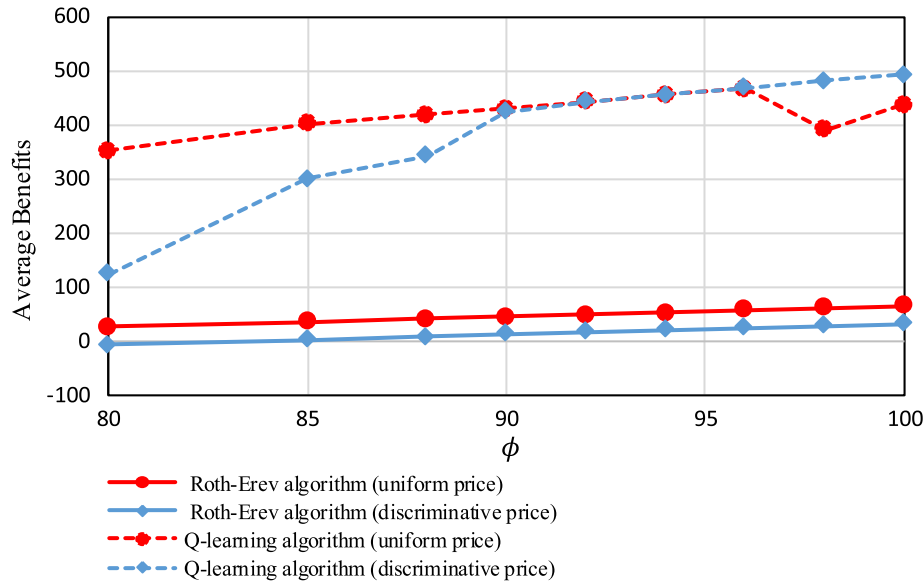
FIGURE 13. Comparison of average firms profits for two algorithms in the two pricing methods.

(1) The bid prices in the uniform price method have higher dispersion, but they are as close to the equilibrium price as possible in the discriminative price method. Our results also show that bid prices in the uniform price method are almost greater than that in the other method.
(2) In the uniform price method, the bidders bid below their maximum volume to decrease the equilibrium price, but in the discriminative price method, with the increase in supply quantity, the bidders bid their maximum volume.
(3) When the supply quantity is low, the government gets more revenue under the discriminative price method, and the firms earn more revenue in the uniform price method, but when the supply quantity is large, the advisable method for the government is the uniform price method and is the other method for the firms.
(4) For low supply quantity, the clearing price and consequently the carbon trading price in the discriminative price are much more than those in the uniform price method, but when supply quantity is abundant, the clearing price and carbon trading price are almost equal in two methods.
(5) The amount of carbon permits traded in the carbon trading market in two pricing methods is different from each other. Because of bid shading, it is more stable in the uniform price than the discriminative price method. With the increase in supply quantity, the tradable carbon permits increase in the discriminative price method. Therefore, when the supply quantity is scarce in the auction, the tradable carbon permits are low for two methods, but when the supply quantity increases, they increase in the discriminative price method and are almost without changing in the other method.
(6) The auction efficiency differs in two pricing methods for different supply quantities. When the supply quantity is low (high), the discriminative price method (the uniform price method) is more efficient than the uniform price method (the discriminative price method).
(7) The amounts of emission reduction and firms' benefit reduction relative to the situation that carbon permit is allocated for free depend on carbon supply quantity. A decrease in carbon supply quantity results in a further increase in these values. When the carbon supply quantity is low, the firms' benefit reduction in the discriminative price method is greater than the uniform price method and *vice versa*. Therefore, if the government wants to decrease the emissions as much as possible and does not want to harm firms economically, the carbon permits should be auctioned in the uniform price method.

Finally, we compare the Q-learning algorithm with the Roth–Erev algorithm. The results demonstrate our method outperforms the Roth–Erev algorithm, and the agents obtain more benefits by using this method.

In summary, the performance of the two pricing methods depends on the amount of carbon permits allocated by the government in the auction market. Incorporating abatement activities and their costs into the presented model can be investigated for future research.

## References

[1] A. Akbari-Dibavar, B. Mohammadi-Ivatloo and K. Zare, Electricity market pricing: uniform pricing vs. pay-as-bid pricing. *Electricity Markets* (2020) 19–35.

[2] L.M. Ausubel, P. Cramton, M. Pycia, M. Rostek and M. Weretka, Demand reduction and inefficiency in multi-unit auctions. *Rev. Econ. Stud.* **81** (2014) 1366–1400.

[3] B. Biaisa and A.M. Faugeron-Crouzetb, IPO Auctions: English, Dutch, . . . French, and Internet. *J. Financ. Intermed.* **11** (2002) 9–36.

[4] C. Böhringer and A. Lange, On the design of optimal grandfathering schemes for emission allowances. *Eur. Econ. Rev.* **49** (2005) 2041–2055.

[5] E. Bonabeau, Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci. U.S.A.* **99** (2002) 7280–7287.

[6] K. Cao, X. Xu, Q. Wu and Q. Zhang, Optimal production and carbon emission reduction level under cap-and-trade and low carbon subsidy policies. *J. Clean. Prod.* **167** (2017) 505–513.

[7] E. Cardona Santos, H. Storm and S. Rasch, The cost-effectiveness of conservation auctions in the presence of asset specificity: an agent-based model. *Land Use Policy* **102** (2021) 104907.

[8] Q. Chai, Z. Xiao, K. Hung Lai and G. Zhou, Can carbon cap and trade mechanism be beneficial for remanufacturing? *Int. J. Prod. Econ.* **203** (2018) 311–321.

[9] F. Chr Matthes and K. Neuhoff, Auctioning in the European Union Emissions Trading Scheme. ko-Institut & University of Cambridge (2007).

[10] S. Clò, Grandfathering, auctioning and carbon leakage: assessing the inconsistencies of the new ETS Directive. *Energy Policy* **38** (2010) 2420–2430.

[11] R.-G. Cong and Y.-M. Wei, Auction design for the allocation of carbon emission allowances: Uniform or discriminatory price? *Int. J. Energy Environ.* **1** (2010) 533–546.

[12] R.G. Cong and Y.M. Wei, Experimental comparison of impact of auction format on carbon allowance market. *Renew. Sustain. Energy Rev.* **16** (2012) 4148–4156.

[13] P. Cramton and S. Kerr, Tradable carbon permits auctions: How and why to auction not grandfather. *Energy Policy* **30** (2002) 333–345.

[14] N.C. Dormady, Carbon auctions, energy markets & market power: an experimental analysis. *Energy Econ.* **44** (2014) 468–482.

[15] S. Du, W. Tang and M. Song, Low-carbon production with low-carbon premium in cap-and-trade regulation. *J. Clean. Prod.* **134** (2016) 652–662.

[16] T. Hattori and S. Takahashi, Discriminatory Versus Uniform Auction: Evidence From JGB Market. Available SSRN. (2020).

[17] J. Hu and M.P. Wellman, Multiagent reinforcement learning theoretical frame work and an algorithm. *ICML* **98** (1998) 242–250.

[18] J. Hu and M.P. Wellman, Nash Q-learning for general-sum stochastic games. *J. Mach. Learn. Res.* **4** (2004) 1039–1069.

[19] M.X. Jiang, D.X. Yang, Z.Y. Chen and P.Y. Nie, Market power in auction and efficiency in emission permits allocation. *J. Environ. Manage.* **183** (2016) 576–584.

[20] D. Matthäus, Designing effective auctions for renewable energy support. *Energy Policy* **142** (2020) 11462.

[21] S. Min Yu, Y. Fan, L. Zhu and W. Eichhammer, Modeling the emission trading scheme from an agent-based perspective: system dynamics emerging from firms' coordination among abatement options. *Eur. J. Oper. Res.* **286** (2020) 1113–1128.

[22] M.J. Poursalimi Jaghargh and H.R. Mashhadi, An analytical approach to estimate structural and behavioral impact of renewable energy power plants on LMP. *Renew. Energy* **163** (2021) 1012–1022.

[23] M. Rahimiyan and H. Rajabi Mashhadi, Supplier's optimal bidding strategy in electricity pay-as-bid auction: comparison of the Q-learning and a model-based approach. *Electr. Power Syst. Res.* **78** (2008) 165–175.

[24] M. Rahimiyan and H. Rajabi Mashhadi, An adaptive Q-Learning algorithm developed foragent-based computational modeling of electricity market. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **40** (2010) 547–556.

[25] S.M. Sadr, H. Rajabi Mashhadi and M.E. Hajiabadi, Evaluation of price-sensitive loads' impacts on LMP and market power using LMP decomposition. *Iran. J. Electr. Electron. Eng.* **12** (2016) 154–167.

[26] T.W. Sandholm and R.H. Crites, On multiagent Q-learning in a semi-competitive domain. In: *International Joint Conference on Artificial Intelligence* (1995) 191–205.

[27] W. Sugiyarto, An Analysis of the Performance of the Indonesian Treasuries Market. Queensland University of Technology (2020).

[28] L. Tang, J. Wu, L. Yu and Q. Bao, Carbon allowance auction design of China's emissions trading scheme: a multi-agent-based approach. *Energy Policy* **102** (2017) 30–40.

[29] G. Tesauro and J.O. Kephart, Pricing in agent economies using multi-agent Q-learning. *Auton. Agent. Multi. Agent. Syst.* **5** (2002) 289–304.

[30] J.J.D. Wang and J.F. Zender, Auctioning divisible goods. *Econ. Theory* **19** (2002) 673–705.

[31] U. Wilensky and W. Rand, An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo. Mit Press (2015).

[32] G. Xiong, T. Hashiyama and S. Okuma, An electricity supplier bidding strategy through Q-Learning. *IEEE Power Eng. Soc. Summer Meeting* **3** (2002) 1516–1521.

[33] G. Xiong, S. Okuma and H. Fujita, Multi-agent based experiments on uniform price and pay-as-bid electricity auction markets. In: *2004 IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies. Proceedings* (2004) 72–76.