Applied Energy 195 (2017) 1000–1011

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

# Competition, risk and learning in electricity markets: An agent-based simulation study

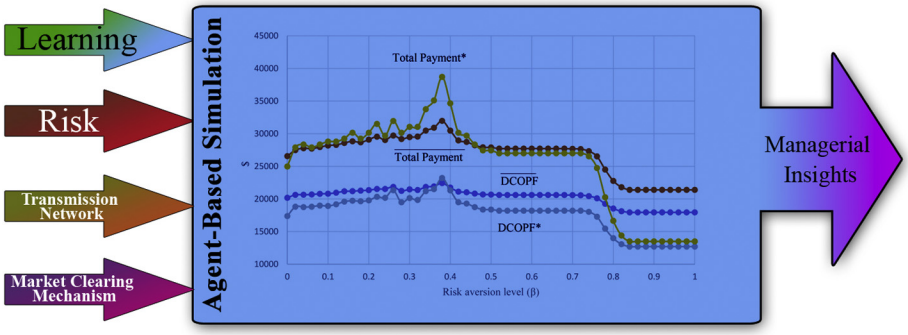
Danial Esmaeili Aliabadi *, Murat Kaya, Guvenc Sahin

*Sabanci University, Faculty of Engineering and Natural Sciences, Istanbul, Turkey*

## HIGHLIGHTS

- The first ABMS study to consider both learning and risk aversion of GenCos.
- Presenting an agent-based simulation of GenCos' behavior in electricity markets.
- Conducting a large-scale analysis with a wide range of learning model parameters.
- Analyzing effects of risk aversion on GenCos' bid prices, profits, and learning.
- Showing that some level of risk aversion can improve GenCos' profits.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Article history:*
Received 21 December 2016
Received in revised form 24 March 2017
Accepted 25 March 2017

*Keywords:*
Electricity markets
Risk aversion
Q-learning
Agent-based simulation
Imperfect competition

## ABSTRACT

This paper studies the effects of learning and risk aversion on generation company (GenCo) bidding behavior in an oligopolistic electricity market. To this end, a flexible agent-based simulation model is developed in which GenCo agents bid prices in each period. Taking transmission grid constraints into account, the ISO solves a DC-OPF problem to determine locational prices and dispatch quantities. Our simulations show how, due to competition and learning, the change in the risk aversion level of even one GenCo can have a significant impact on all GenCo bids and profits. In particular, some level of risk aversion is observed to be beneficial to GenCos, whereas excessive risk aversion degrades profits by causing intense price competition. Our comprehensive study on the effects of Q-learning parameters finds the level of exploration to have a large impact on the outcome. The results of this paper can help GenCos develop bidding strategies that consider their rivals' as well as their own learning behavior and risk aversion levels. Likewise, the results can help regulators in designing market rules that take realistic GenCo behavior into account.



## 1. Introduction

In this work, we present a wholesale electricity market simulation with learning agents. The agents are power generation companies (GenCos) that engage in repetitive hourly pool trading. We are concerned with how the learning behavior and risk aversion of competing GenCos will shape GenCo bid prices and profit levels.

Electricity markets are oligopolies, and electricity demand is often considered inelastic in the short term with respect to price. In addition, transmission line constraints, and the relative locations of electricity demand and supply can provide market power to individual GenCos. Due to all these reasons, GenCos can bid above their marginal costs and obtain positive profit. This possibility, and the importance of the power sector for the economy has triggered a

* Corresponding author.
*E-mail addresses:* danialesm@sabanciuniv.edu (D. Esmaeili Aliabadi), mkaya@sabanciuniv.edu (M. Kaya), guvencs@sabanciuniv.edu (G. Sahin).

http://dx.doi.org/10.1016/j.apenergy.2017.03.121
0306-2619/

**Nomenclature**

*Indices*

| | |
|---|---|
| $i, k, l$ | GenCos or nodes |
| $j$ | bids |
| $t$ | simulation iterations |

*Sets*

| | |
|---|---|
| $BR$ | set of transmission lines |
| $B_i$ | set of available bids of GenCo-$i$ |
| $H_{ij}$ | history of all realized profits for submitting bid price alternative $b_{ij}$ |

*Parameters*

| | |
|---|---|
| $n$ | number of nodes in the network |
| $P_i^{max}$ | the maximum generation capacity of GenCo-$i$ (MW) |
| $C_i$ | production cost of GenCo-$i$ ($/MW h) |
| $D_k$ | power demand at node-$k$ (MW) |
| $F_{kl}^{max}$ | thermal limit for real power flow on line $kl$ |
| $y_{kl}$ | negative of the susceptance value for line $kl$ |
| $\beta_i$ | risk aversion level of GenCo-$i$ |
| $\alpha_{it}$ | recency rate of GenCo-$i$ at iteration $t$ |
| $\alpha_{i0}$ | initial recency rate of GenCo-$i$ |
| $\epsilon_{it}$ | exploration parameter of GenCo-$i$ at iteration $t$ |
| $\epsilon_{i0}$ | initial exploration parameter of GenCo-$i$ |
| $max_t$ | number of iterations |

*Variables*

| | |
|---|---|
| $\Theta_i$ | voltage angle at node-$i$ *(radians)* |
| $LMP_i$ | locational marginal price at node-$i$ ($/MW h) |
| $P_i$ | power injected by GenCo-$i$ (MW) |
| $b_{ij}$ | $j$th bid price alternative of GenCo-$i$ ($/MW h) |
| $b_i$ | simplified notation for $b_{ij}$ ($/MW h) |
| $\underline{b_i^*}$ | the best identified bid price of GenCo-$i$ ($/MW h) |
| $\overline{b_i}$ | average bid price submitted by GenCo-$i$ over iterations ($/MW h) |
| $r_{ij}$ | realized profit of GenCo-$i$ for submitting $b_{ij}$ ($) |
| $r_i$ | simplified notation for $r_{ij}$ ($) |
| $r_i^*$ | profit of the best identified bid price of GenCo-$i$ ($) |
| $\overline{r_i}$ | average realized profit of GenCo-$i$ over iterations ($) |
| $Q_{ij}$ | $Q$-value of GenCo-$i$ for submitting $b_{ij}$ ($) |
| $Q_{ij}^r$ | risk-modified $Q$-value ($) |
| $\overline{Q_i}$ | average $Q$-value of GenCo-$i$ over different scenarios ($) |
| $CP_i$ | cumulative profit of GenCo-$i$ ($) |
| $\overline{CP_i}$ | average cumulative profit of GenCo-$i$ over different scenarios ($) |

wave of research into the strategic bidding behavior of GenCos (see, for example, [1–4]).

In this study, building on [5], we develop a flexible agent-based simulation model (ABMS) to characterize the evolution of a dynamic electricity market under transmission grid constraints. ABMS models have gained popularity in electricity market research because they offer advantages over game-theoretical models such as the ability to model heterogeneous players and observe the dynamic evolution of the market. The distinction of our ABMS model is that it considers both the *learning behavior* and *risk aversion* of GenCos.

In practice, GenCos often make bid decisions without proper information on the characteristics (such as capacity, cost, and/or financial situation) and bid history of competing GenCos. There is, however, also the potential for learning due to the repetitive nature of trading as GenCos interact with each other every day and gain experience. Due to learning and adaptation, GenCos can be expected to exhibit time-variant bidding policies. While learning individually, through its bids, each GenCo also has an impact on all the prices and dispatch quantities in the market. In this study, we are interested in observing how the collective learning of the GenCos will change the market. For example, will a GenCo discover its strategic advantage, such as low cost or a favorable position in the network, and learn over time to take advantage of it? We model GenCo learning using the $Q$-learning approach. In particular, we extend the learning model of [5,6] by considering time-dependent learning model parameters, similar to [7].

The literature that addresses GenCo behavior generally assumes risk-neutral decision makers whose objective is to maximize only the expected profit. In reality, however, GenCos may act risk-averse because they are exposed to increased levels of risk due to fluctuations in hourly prices and dispatched power quantities. To study the effects of risk aversion, we adopt a model in which risk is captured through the variance of past realized profits.

The major contributions of this paper can be summarized as follows: First, this is the first ABMS paper that studies the joint effects of learning behavior and risk aversion on GenCo bid prices and profits. Second, we present a flexible simulation model that can characterize the evolution of a dynamic electricity market under transmission constraints and time-dependent learning parameters. Third, we show that a certain level of risk aversion can improve GenCos' profits, whereas excessive risk aversion decreases profits due to intense price competition. Finally, we present a comprehensive study on the effects of $Q$-learning parameters, in which we find the level of exploration to have a large impact on results.

The remainder of the paper is organized as follows: In Section 2 we summarize the related literature. In Section 3, we present the model with risk-neutral GenCos, discussing the network and market structures as well as our learning model. In Section 4, we illustrate the learning model and simulation algorithm through two case studies. In Sections 5 and 6, we discuss the model with risk-averse GenCos and the related simulation study, respectively. Section 7 presents the simulation study regarding the effects of $Q$-learning parameters. Finally, we discuss the implications of our results in Section 8, and conclude in Section 9.

## 2. Literature survey

Ventosa et al. [8] provide a review of electricity market modeling approaches, classifying the literature into optimization, equilibrium and simulation models. Among these, game-theoretical models aim to characterize the equilibrium when players compete in quantity (Cournot competition [9]), in price (Bertrand competition [10]), or by submitting supply functions (supply function equilibrium [2,11]). There are also the more general conjectural variation type models [12–14].

Game-theoretical models have been extensively used because they offer insights into the strategic behavior of players and allow an easy derivation of equilibrium results. However, they are too stylized to reflect the realities of complex electricity markets [1]. Almost all game-theoretical models assume players to be rational, which often does not hold in practice, and implicitly assume GenCo behavior not to change over time. In addition, transmission grid constraints are ignored in most game-theoretic electricity market

studies [2,15,16], despite the fact that physical network characteristics can lead to important differences in results [17].

Due to these shortcomings of game-theoretical models, Agent-Based Modeling and Simulation has become a popular choice in the modeling of electricity markets [18–21]. ABMS offers a flexible computational approach to model GenCo agents that are heterogeneous in parameters such as generation cost and capacity, location in the network, and possessed information. Moreover, ABMS imposes minimal information requirements and avoids the multiple equilibrium issues of the game-theoretical models. ABMS allows the modeling of the two-way interaction between agents and the market. The evolution of the resulting models can be observed in detail, even for dynamic systems that are not in equilibrium. Agent-based approach also assists in modeling learning and adaptation in a dynamic environment.

In the ABMS electricity market literature, different types of reinforcement learning methods have been used to simulate the intelligence of the players [21]. The most popular of these include Erev-Roth reinforcement learning [22], Q-learning [23], and the temporal difference algorithm [24]. In [25,26], modified versions of the Erev-Roth reinforcement approach has been applied to electricity markets.

Q-learning is a model-free and state-dependent algorithm that was originally designed to be used with a Markov Decision Process. Q-learning algorithms have been extensively used in many applications such as industrial control, time sequence prediction and robot soccer competition [27]. In [28], GenCos learn through a Q-learning algorithm, yet when bidding for capacity, they also consider their competitors' actions through a conjectural-variation based strategy. In [29], the effect of market power mitigation strategies is analyzed through an agent-based study with Q-learning. In the current paper, we apply the Q-learning model used in [7], which extends the learning model of [5] by considering time-decaying parameters. The use of time-decaying parameters is akin to the Metropolis criterion in simulated annealing (i.e., the SA-Q-learning algorithm in [30,28]). Similar to [6], we assume a state-independent version of the method in which Q-values are expressed as functions of actions only.

Dahlgren et al. [31] provide an early review of risk assessment methods in energy trading. In [32], the contract quantity determination problem is considered under uncertain generation and imbalance prices. A number of researchers have formulated stochastic programming models to develop bidding strategies under supply and price risks [33–36]. Zheng et al. [37] provide a recent review of the stochastic optimization literature that address the unit commitment problem. GenCos' self-scheduling problem is also widely studied [38–44].

The aforementioned papers assume price-taking GenCos operating under perfect competition. Ventosa et al. [8]'s survey cites [45] as the only work that addresses the risk management problem of GenCos under imperfect competition, in which case GenCos become price-makers. In [45], a Monte Carlo simulation model is developed to capture hydro production and demand risks in electricity markets under Cournot competition. The authors use risk measures such as value-at-risk (VaR) and profit-at-risk (PaR). In [46], similar to our study, a competitive market with network constraints is considered. The authors solve a bi-level optimization problem in which competing GenCos submit linear supply functions at the first stage, and the ISO determines the dispatch at the second stage. In [47], both pool and bilateral-contract structures are analyzed. In [48], the integrated risk management problem of a hydrothermal GenCo in an oligopolistic market is considered. The risk exposure due to fuel price, water inflow, electricity price and power demand uncertainties are represented by the conditional value-at-risk (CVaR) approach. In [49], bidding strategies for a single price-taker hydro GenCo are studied. Uncertainty about competitor GenCo offers are represented through the residual demand curve. In [50], the effect of risk aversion and forward markets on capacity expansion and forward hedging decisions of GenCos are analyzed. Similar to our work, these authors observe that due to competition, GenCos in a Cournot duopoly may obtain higher expected profits as they become more risk averse. A similar observation is made in [51] regarding the investment decisions of natural gas suppliers in an oligopolistic gas market.

Another stream of risk-related papers are those that address the generation portfolio selection problem of a single GenCo. For instance, [52] presents a stochastic programming model for the integrated portfolio selection and scheduling problem of a risk-averse hydro producer. In [53], a Monte-Carlo simulation tool is developed to optimize a power portfolio composed of physical and financial assets. In [54], a detailed production model with natural gas, wind and cascaded hydro units is considered. Using agent-based simulation and Monte Carlo approaches, [55] discusses the effect of risk aversion on power plant investment decisions. In [56], a Monte Carlo simulation is developed for assessment of low-carbon power plant proposals.

As discussed in the paragraphs above, the learning behavior and risk attitude of GenCos have been separately studied in the literature. Yet, their joint effect has not so far been investigated, which presents a gap in the literature. We are aware of only two pieces of work that address both learning behavior and risk attitudes; however, their model structures as well as definitions of learning and risk are different from our work. Liu and Wu [57] present a stochastic optimal control mechanism in an oligopolistic market in which GenCos engage in Cournot competition. The adaptation mechanism in this model is somewhat similar to the learning in our model. The [57] model, however, ignores the bidding and price formation details in the market, and the network physical structure. Rahimiyan and Mashhadi [58] consider both Q-learning (a fuzzy version) and risk attitude of GenCos. However, different from our model and the literature in general, GenCo risk attitude is characterized as a combination of certain Q-learning parameter values, without using a separate risk component in the model.

## 3. Model with risk-neutral GenCos

In this section, we assume risk-neutral GenCos that aim to maximize expected profit. We first discuss the network representation and market structure in the framework of our study. Then, we provide the details of the GenCo learning model.

### 3.1. The network and market structure

We consider only the day-ahead market, ignoring the futures markets and real-time markets. The market is an oligopoly with a relatively small number of GenCos each having a single production unit. All parameters related to GenCos, including demand, capacity and costs, are steady. Line or generation outages are ignored.

The physical transmission grid is represented using a network in which nodes correspond to GenCos and arcs correspond to transmission lines between GenCos. The GenCo that connects to the system at node-$i$ is referred to as GenCo-$i$. GenCo-$i$ has generation capacity $P_i^{max} > 0$ and marginal production cost $C_i$. Power demand (load) in node-$i$, $D_i$, is assumed to be constant and price-inelastic. The transmission line between nodes $k$ and $l$ has capacity $F_{kl}^{max}$ and susceptance $-y_{kl}$.

For every period, corresponding to an hour in the day-ahead market, each GenCo submits a bid composed of a power quantity and price:

- **Bid quantity:** The GenCo is assumed to bid all its production capacity, without capacity withholding. This assumption implies that systems with storage capacity are ignored.
- **Bid price:** The GenCo chooses a bid price $b_i$ ($/MW h) among the exogenously-given bid price alternatives ($b_{ij} \in B_i$). The bid price alternatives range from the GenCo's marginal production cost $C_i$ to a given price cap in the market.

The information set of each GenCo-$i$ consists of the history $H_{ij}$ of its own realized profit values for each bid price alternative $b_{ij}$. The GenCo has no information about the generation capacity, marginal cost, bid prices or profits of other GenCos, or the total number of GenCos in the system.

The market is run by the Independent System Operator (ISO), which collects the bids and clears the market. For each period, the ISO solves the following DC-OPF problem as an approximation to the underlying AC-OPF problem. Note that AC-OPF problems are typically approximated by the more tractable DC-OPF problems that consider linearized power constraints [59].

$$min \quad \sum_{i=1}^{n} b_i P_i \tag{1}$$

$$\text{subject to} \quad P_k - D_k = \sum_{(k,l)\in BR} y_{kl}(\theta_k - \theta_l), \quad \forall k \in \{1,\ldots,n\}, \tag{2}$$

$$P_i \leqslant P_i^{max}, \quad \forall i \in \{1,\ldots,n\}, \tag{3}$$

$$|y_{kl}(\theta_k - \theta_l)| \leqslant F_{kl}^{max}, \quad \forall (k,l) \in BR. \tag{4}$$

The objective (1) is to minimize the system-wide cost of power supply. Constraint (2) allows the surplus power in each node to flow via the transmission lines to the connected nodes. Constraint (3) is the generation capacity constraint of each GenCo. Constraint (4) is the power flow constraint on each transmission line. The problem (1)–(4) is a linear optimization problem as we assume a DC representation of the transmission network.

By solving the problem, the ISO determines the power $P_i$ to be dispatched by each GenCo-$i$, the voltage angle $\theta_i$ at each node-$i$, and the Locational Marginal Price $LMP_i$ at each node-$i$, which is the shadow price for Constraint (2). LMP at a node corresponds to the minimum cost of fulfilling the demand for an additional unit (MW) of power at that particular node. Based on the solution, each GenCo is paid the LMP at its location node multiplied by its power dispatch. Thus, the profit $r_{ij}$ of GenCo-$i$ from bidding price $b_{ij}$ at a particular period becomes

$$r_{ij} = P_i(LMP_i - C_i). \tag{5}$$

Note that because $P_i$ and $LMP_i$ values are determined as a function of all bids submitted to the ISO, each GenCo's profit is affected by the bid price choices of all GenCos. We refer to GenCo-$i$'s profit simply by $r_i$ when the particular bid price $b_{ij}$ that resulted in the profit is not relevant for the discussion.

### 3.2. The learning model

To model the learning behavior of GenCos, we use the modified $Q$-learning algorithm of [7] in which each GenCo (agent) experiments with bid price alternatives and learns through experience. GenCo-$i$ keeps a set $H_{ij}$ of past realized profits from bidding the price alternative $b_{ij}$. This includes the zero profit realizations due to rejected bids. For each price alternative $b_{ij}$, the GenCo calculates the $Q$-value $Q_{ij}$ which denotes the weighted average of past realized profits from bidding $b_{ij}$. $Q_{ij}$ captures the expected profit GenCo-$i$ believes to obtain by bidding this price in the subsequent period.

When GenCo-$i$ bids price $b_{ij}$ and obtains the profit $r_{ij}$ in a particular period, the relevant $Q$-value is updated as

$$Q_{ij} = (1 - \alpha_{it})Q_{ij} + \alpha_{it} r_{ij}. \tag{6}$$

The history set is also updated as $H_{ij} \leftarrow H_{ij} \cup r_{ij}$. The other bid price alternatives' $Q$-values are unchanged. In this equation, the *recency rate* $\alpha_{it}$ determines the weight given to the most recent profit observation. If $\alpha_{it} = 1$, the last obtained profit ($r_{ij}$) becomes the $Q$-value. In this case, the agent uses only the last period information for that price alternative, ignoring history. At the other extreme, if $\alpha_{it} = 0$, the $Q$-value will not be updated. To facilitate convergence, we assume $\alpha_{it}$ to start at $\alpha_{i0}$, and decay linearly over periods to $\alpha_{i0}/10$ according to the equation $\alpha_{it} = (1 - t/max_t)(\alpha_{i0}) + (\alpha_{i0}/10)(t/max_t)$, where $max_t$ is the number of periods. We have also studied the case of exponentially decaying $\alpha_{it}$ and found no significant difference in results.

The bid price alternative that maximizes the expected profit in a particular period is labeled as the GenCo's *best identified bid price*. Note that the best identified bid price values are by definition time-dependent. We ignore their time index because the meaning will be clear from the context.

$$b_i^* = \underset{b_{ij}}{Max}\ Q_{ij}. \tag{7}$$

In choosing its bid price, the GenCo uses an $\epsilon$-greedy action selection rule [60], which is characterized by the *exploration parameter* $\epsilon$: In each period, GenCo-$i$ submits its best identified bid price $b_i^*$ with probability $1 - \epsilon_{it}$. With probability $\epsilon_{it}$, the GenCo submits a randomly chosen price $b_{ij}$ from its set of bid price alternatives. Each alternative has an equal probability of being chosen. This randomization helps the GenCo in assessing the performance of different bid price alternatives. The approach aims to strike a balance between exploitation of the best identified bid price and exploration of possibly better bid prices. A high exploration parameter $\epsilon$ would cause the GenCo to search for better bid prices most of the time, slowing learning. A low $\epsilon$, on the other hand, may lead to local optimum solutions by causing the GenCo to stick prematurely to a particular $b_i^*$.

We use a time-decaying exploration parameter. This approach, which is similar to [30], is different from most works in the literature, in which $\epsilon$ is fixed (e.g., [6]). Starting from a relatively high initial value of $\epsilon_{i0}$, the parameter decreases over time towards zero according to the equation $\epsilon_{it} = max\{0, \epsilon_{i0} + 8t(\epsilon_{i0} - 1)/max_t\}$. Hence, exploration is favored in the initial periods. Over the course of the simulation, GenCo's experience about the profitability of different bid price alternatives builds up, decreasing the need for exploration. Thus, the GenCo is more likely to exploit its experience by submitting its $b_i^*$. Because $\epsilon$ decreases over time to zero, the GenCo's bid price choice will converge to an alternative that hopefully maximizes its Q function.

The aforementioned learning model captures the dynamics of a GenCo's bidding behavior over time. We refer to a GenCo that bids to maximize its utility and is subject to learning through this model as a *learning GenCo*. Note that in the model, the GenCo does not take any strategic action, that is, the GenCo does not consider the actions of other GenCos explicitly in its decision process. In fact, it does not have information on other GenCos. The GenCo is modeled as a simple agent that learns only from its own experience. GenCos' collective behavior, however, may lead to strategic outcomes.

## 4. Simulation study with risk-neutral GenCos

A simulation run in our study consists of $max_t$ iterations. Each iteration corresponds to the settlement of an hourly auction (one period) in the day ahead market. In each iteration, GenCos simultaneously bid prices $b_i$ to the ISO. The ISO then solves the DC-OPF

problem given in (1)–(4) to determine the power $P_i$ to be dispatched by each GenCo-$i$, and the nodal price $LMP_i$ at each node.

At the beginning of the simulation, the following variables are initialized as $t = 1, Q_{ij} = 0, H_{ij} = \varnothing$, whereas $\alpha_{i0}$ and $\epsilon_{i0}$ are set to their initial values. One simulation run with 2000 iterations takes less than two seconds on an Intel Core i7 @ 3.2 GHz computer with 24 GB RAM. To obtain robust results, we report the average result over a number of simulation runs, each having a different random number seed. This is similar to the literature [15]. The random numbers are used for simulating the $\epsilon$-greedy action selection rule of each GenCo in each iteration: First, in determining whether the GenCo submits its best identified bid price, and if this price is not to be submitted, in determining the bid price to submit among the alternatives.

To illustrate our learning model and simulation algorithm, we use two case studies that are based on [6]. In both case studies, we consider the five-node transmission grid presented in Fig. 1. This network topology, which follows from [5], is inspired by the real Pennsylvania-NewJersey-Maryland (PJM) five node power system. The (negative) susceptance and the maximum flow values of the transmission lines are summarized in Table 1. Node-3 is the reference bus in this system.

### 4.1. Case 1: Two learning GenCos and a unique Nash equilibrium

In this case, GenCo-1 and GenCo-5 are the learning GenCos, both having $\epsilon_{i0} = 0.9$ and $\alpha_{i0} = 0.1$. Other GenCo parameters are summarized in Table 2. We report results from a single sample simulation run that has 300 iterations.

Table 3 shows the profits resulting from each possible bid price profile $(b_1, b_5)$ for the learning GenCos. The highlighted profile (20, 50) is the only pure strategy Nash equilibrium of the stage game. This profile also happens to provide the maximum total profit for GenCos. Fig. 2 illustrates the evolution of $Q$-values for each bid price alternative during the simulation. We observe that the two learning GenCos eventually reach the Nash equilibrium with bids ($b_1 = 20$ and $b_5 = 50$). The graph for GenCo-5 shows that it takes some iterations to learn to bid \$50/MW h.

### 4.2. Case 2: Three learning GenCos and multiple Nash equilibria

In this case, GenCo-2 in Case 1 also becomes a learning GenCo, and its bid price alternatives are extended from $\{20\}$ to $\{20, 30, 40, 50\}$. With this new setting, as seen in the profit values of Table 4, the stage game has two Nash equilibria as $\{20, 40, 50\}$ and $\{30, 50, 50\}$. The initial $Q$-learning parameters are $\epsilon_{i0} = 0.85$ and $\alpha_{i0} = 0.15$ for all GenCos.

We aim to observe where the simulations will converge. To this end, 10,000 simulation runs are conducted, each having 2000 iterations. Overall, we observe most simulation runs to converge to either one of the two Nash equilibria, or a state that provides a similar profit profile to a Nash equilibria. In fact, 65.6% of the runs converge to the Nash equilibrium $\{20, 40, 50\}$ and 6.25% converge to the Nash equilibrium $\{30, 50, 50\}$. Compared to the latter, the former equilibrium provides higher social welfare, i.e., lower cost of power as measured by the DC-OPF objective function value, and a more equitable profit distribution among GenCos.

The simulation runs also converge to state $\{30, 40, 50\}$ with probability 25% and to $\{40, 50, 50\}$ with probability 3.12%. These Non-Nash states provide an identical profit profile to one of the Nash equilibria. Such states are referred to as *semi-Nash* by [7]. Apparently, in search of better profits, the $Q$-learning behavior of GenCos can make the market converge to even a non-Nash state if this state provides reasonable profits.

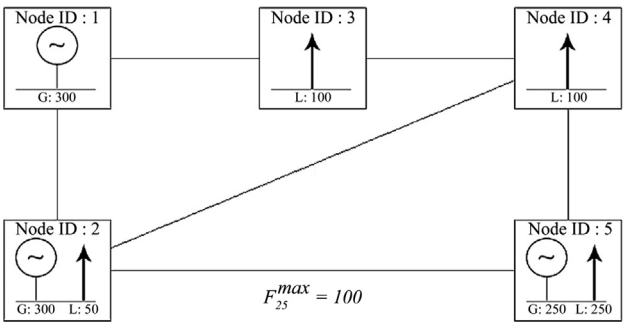



**Fig. 1.** The transmission grid in case studies.

**Table 1**
Transmission line parameters in Case 1.

| k/l | $y_{kl}$ | $F_{kl}^{max}$ (MW) |
|---|---|---|
| {1/2, 1/3, 2/4, 3/4, 4/5} | 4 | No limit |
| 2/5 | 4 | 100 |

**Table 2**
GenCo parameters in Case 1.

| ID | $P_i^{max}$ (MW) | $C_i$ (\$/MW h) | $B_i$ (\$/MW h) |
|---|---|---|---|
| 1 | 300 | 20 | {20, 30, 40, 50} |
| 2 | 300 | 20 | {20} |
| 5 | 250 | 30 | {30, 40, 50} |

**Table 3**
Profits $(r_1, r_5)$ obtained from possible bid price profiles $(b_1, b_5)$ in Case 1.

| | $b_5$ | | |
|---|---|---|---|
| $b_1$ | 30 | 40 | 50 |
| 20 | (428.57, 0) | (857.14, 1214.29) | **(1285.71, 2428.57)** |
| 30 | (0, 0) | (416.67, 1583.33) | (416.67, 3166.67) |
| 40 | (0, 0) | (0, 2000) | (833.33, 3166.67) |
| 50 | (0, 0) | (0, 2000) | (0, 4000) |

## 5. Model with risk-averse GenCos

Here, we present a model with risk averse GenCos, which encompasses the risk-neutral model of Section 3 as a special case. In this risk-averse model, the utility of bid price alternative $b_{ij}$ to GenCo-$i$ is increasing in its $Q$-value $Q_{ij}$, and decreasing in the standard deviation of past realized profits from this alternative, which are recorded in the set $H_{ij}$. Accordingly, GenCo-$i$'s best identified bid price is determined using the *risk-modified* $Q^r$-values as

$$b_i^* = \underset{b_{ij}}{Max}\ Q_{ij}^r \tag{8}$$

$$\text{where} \quad Q_{ij}^r = (1-\beta_i)Q_{ij} - \beta_i \sqrt{\frac{\sum_{r_{ij} \in H_{ij}} (r_{ij} - Q_{ij})^2}{|H_{ij}| - 1}}. \tag{9}$$

Parameter $\beta \in \{0, 1\}$ denotes the *risk aversion level* of the GenCo where $\beta = 0$ corresponds to the risk-neutral case and higher $\beta$ values correspond to more risk-aversion. $Q$-values are updated similar to the risk-neutral model, based on realized profits as given in Eq. (6).

## 6. Simulation study with risk-averse GenCos

To address risk aversion, two modifications are made in the simulation algorithm. First, $b_i^*$ is now determined based on the

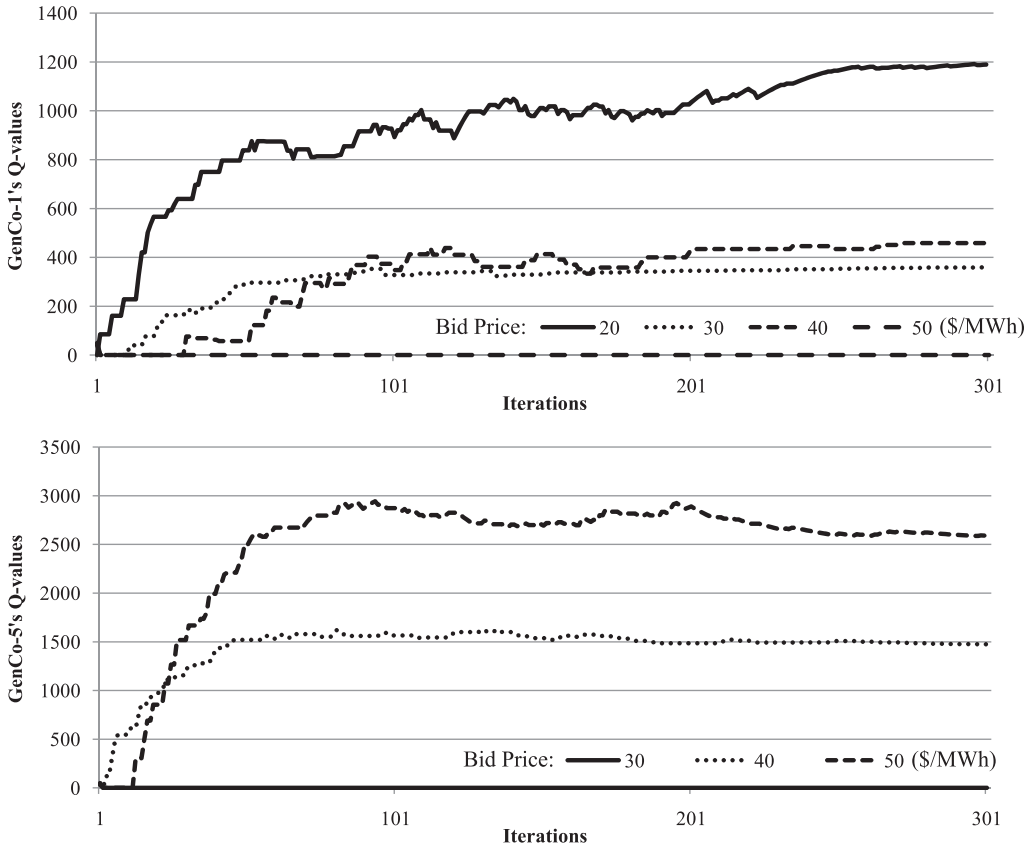



**Fig. 2.** *Q*-value evolutions for GenCo-1 and GenCo-5 in Case 1.

**Table 4**
Profits $\{r_1, r_2, r_5\}$ of bid profiles $\{b_1, b_2, b_5\}$ in Case 2 where Rows: $B_1$, Columns: $B_2$ and separated tables: $B_5$.

| | 20 | 30 | 40 | 50 |
|---|---|---|---|---|
| $b_5 = 30$ | | | | |
| 20 | (428.57, 0, 0) | (3000, 785.71, 0) | (3000, 0, 0) | (3000, 0, 0) |
| 30 | (0, 3000, 0) | (0, 3000, 0) | (2500, 0, 0) | (2500, 0, 0) |
| 40 | (0, 3000, 0) | (0, 3000, 0) | (0, 5000, 2500) | (5000, 0, 2500) |
| 50 | (0, 3000, 0) | (0, 3000, 0) | (0, 5000, 2500) | (0, 7500, 5000) |
| $b_5 = 40$ | | | | |
| 20 | (857.14, 0, 1214.29) | (3428.57, 785.71, 1214.29) | (6000, 1571.43, 1214.29) | (6000, 0, 2000) |
| 30 | (416.67, 2500, 1583.33) | (3428.57, 785.71, 1214.29) | (6000, 1571.43, 1214.29) | (6000, 0, 2000) |
| 40 | (0, 6000, 2000) | (0, 6000, 2000) | (0, 6000, 2000) | (5000, 0, 2500) |
| 50 | (0, 6000, 2000) | (0, 6000, 2000) | (0, 6000, 2000) | (0, 7500, 5000) |
| $b_5 = 50$ | | | | |
| 20 | (1285.71, 0, 2428.57) | (3857.14, 785.71, 2428.57) | **(6428.57, 1571.43, 2428.57)** | (9000, 0, 4000) |
| 30 | (416.67, 2000, 3166.67) | (3857.14, 785.71, 2428.57) | (6428.57, 1571.43, 2428.57) | **(9000, 2357.14, 2428.57)** |
| 40 | (833.33, 5500, 3166.67) | (833.33, 5500, 3166.67) | (6428.57, 1571.43, 2428.57) | (9000, 2357.14, 2428.57) |
| 50 | (0, 9000, 4000) | (0, 9000, 4000) | (0, 9000, 4000) | (0, 9000, 4000) |

risk-modified $Q^r$-values as shown in Eq. (8). Second, the variance component of Eq. (8) is ignored in the initial half of the iterations because there may be too few observations to support the required variance calculations.

We study the effects of risk aversion on GenCos' bid prices and profits using a new case study, Case 3. The grid structure and GenCos' parameters for this case are presented in Fig. 3 and Table 5, respectively. This structure provides GenCo-3 advantage due to zero generation cost, whereas GenCo-4 is at an unfavorable position with a relatively high generation cost. All reported results are averages over 30 random runs each having 2000 iterations. The initial *Q*-learning parameters are $\epsilon_{i0} = 0.85$ and $\alpha_{i0} = 0.15$ for all GenCos.

### 6.1. Identical risk aversion level

Here, we assume all three GenCos to have the same $\beta$. We initially focus on the picture at the end of the simulation. Fig. 4(a) provides the $b_i^*$ value at iteration 2000 for each GenCo, averaged

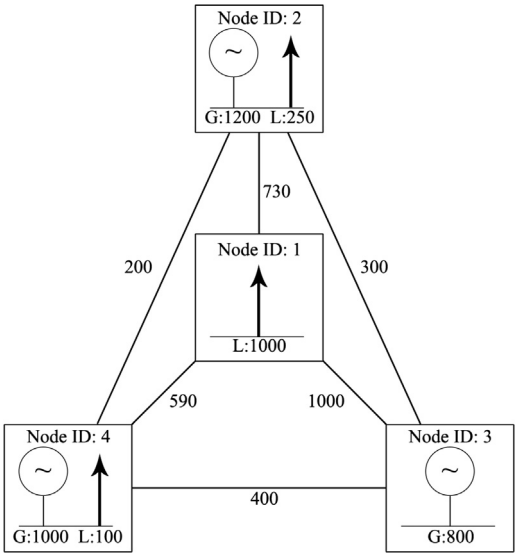



**Fig. 3.** Transmission grid in Case 3.

**Table 5**
GenCo parameters in Case 3.

| ID | $P_i^{max}$ (MW) | $C_i$ ($/MW h) | $B_i$ ($/MW h) |
|---|---|---|---|
| 2 | 1200 | 10 | {10, 20, 30, 40} |
| 3 | 800 | 0 | {9, 18, 20} |
| 4 | 1000 | 15 | {15, 25, 35, 45} |

over 30 runs, as a function of the identical $\beta$. Fig. 4(b) presents the corresponding profit values.

We observe $b_3^*$ and $b_4^*$ to be quite stable for different risk aversion levels. GenCo-3 mostly bids 9. It ventures into bidding 18 only for relatively small $\beta$ values. GenCo-4 bids its maximum price of 45 for any $\beta$ value except zero. GenCo-2, on the other hand, responds to different levels of risk aversion. In fact, the profit results observed in Fig. 4(a) are driven by the changes in $b_2^*$. As $\beta$ increases from 0.00 to around 0.38, $b_2^*$ increases. When risk-neutral, GenCo-2 usually bids 20, but as it becomes risk-averse, this GenCo tries higher bid prices such as 30 or 40 more frequently. These higher bid prices lead to higher profits not only for GenCo-2 itself, but also for its competitor GenCo-3 as well. In fact, both GenCos' individual profits, and also the total profit of all GenCos are maximized at $\beta = 0.38$. Thus, some level of risk aversion in the market could benefit all GenCos.

After reaching a maximum around $\beta = 0.38, b_2^*$ decreases for higher risk aversion levels. In fact, for $\beta \in [0.74, 0.82]$, GenCo-2 becomes excessively concerned about the variability in profits and bids its marginal cost 10 more frequently. For even higher $\beta$ values, GenCo-2 only bids 10, resulting in zero profits. Such low bids by GenCo-2 causes a significant reduction in the profit of competitor GenCo-3 as well. For sufficiently high $\beta$ values, both GenCo-2 and GenCo-3 submit their marginal generation costs to minimize the variability in their profits. GenCo-4, meanwhile, is observed to obtain zero profit at the end of the simulation independent of $\beta$.

We have discussed the end-of-simulation results when each GenCo-$i$ bids its $b_i^*$ as of iteration 2000. While these converged results are of interest, they do not necessarily represent what has happened throughout the simulation, especially during the initial iterations in which most of the learning takes place. Figs. 5(a) and (b) provide the average results over all 2000 iterations, again averaged over 30 runs. A comparison between Figs. 4 and 5 illustrate the effects of GenCo learning and competition over time.

The similarities in shapes indicate strong convergence in bid prices. The differences in bid prices point to changes in bidding behavior over time. In particular, the effect of risk aversion on $b_2^*$ becomes sharper over iterations. GenCo-3 bids higher prices than 9, and GenCo-4 bids lower prices than 45 throughout the iterations. Accordingly, GenCos' profits converge to more extreme levels at the end of the simulation. For $\beta < 0.74$, the competing GenCos, GenCo-2 and GenCo-3, achieve higher profits at the end of the simulation than they do in the initial iterations. For higher $\beta$ values, however, extreme risk aversion of GenCo-2 causes a reduction in both GenCos' profits. Meanwhile, as expected, GenCo-4's profits converge to zero over iterations. All these observations underscore the importance of risk aversion on GenCos' bidding behavior and resulting profit levels in an environment shaped by dynamic learning and competition.

Fig. 6 presents the corresponding DC-OPF optimal value, $\sum b_i P_i$ and the total payment to GenCos, $\sum(LMP_i \times P_i)$. Comparing the end-of-simulation and simulation-average results, we make the following two observations:

- **DC-OPF optimal value:** For almost all $\beta$, the end-of-simulation DC-OPF optimal value is lower than the simulation-average value. Hence, the ISO's auction mechanism seems to be successful in driving GenCos' bid prices down throughout the simulation.
- **Total payment to GenCos:** For $\beta < 0.45$, the total payment to GenCos, hence, their total profit is higher at the end of the simulation than the simulation-average payment. When the GenCos are not very risk-averse, they collectively learn to obtain better profits over time. For $\beta > 0.45$, however, the observation is reversed; risk-averse behavior of GenCo-2 causes a reduction in total GenCo profits. This reduction is especially acute for $\beta > 0.74$.

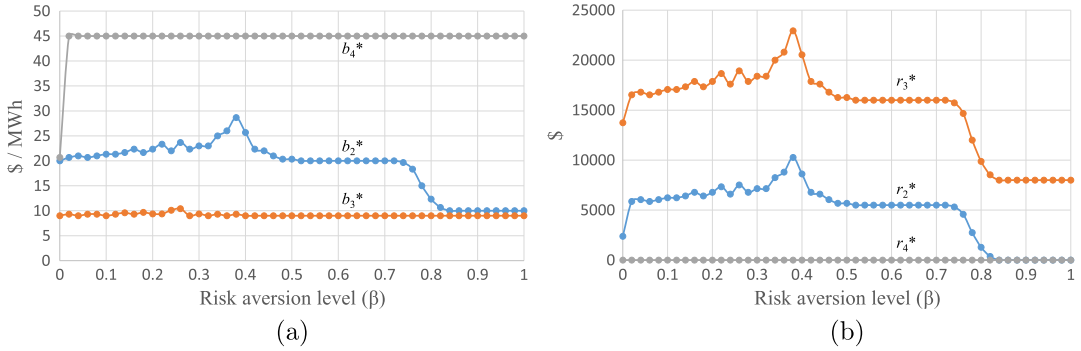



**Fig. 4.** End-of-simulation results. (a) The best identified bid prices. (b) Profits.

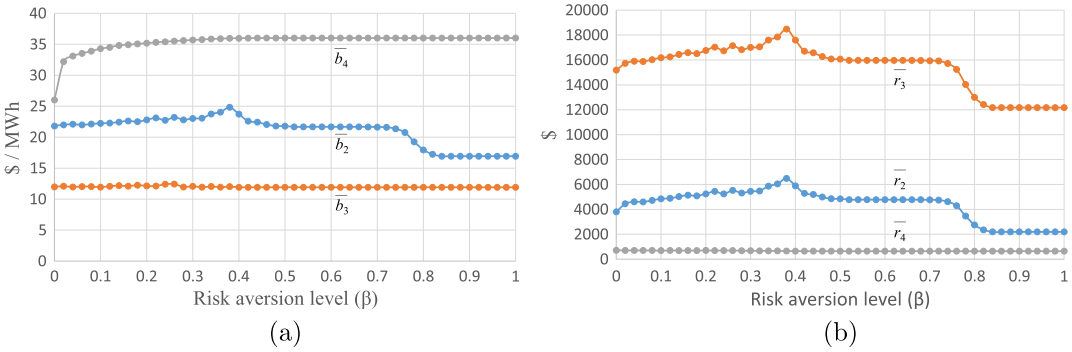



**Fig. 5.** Simulation averages. (a) The best identified bid prices. (b) Profits.

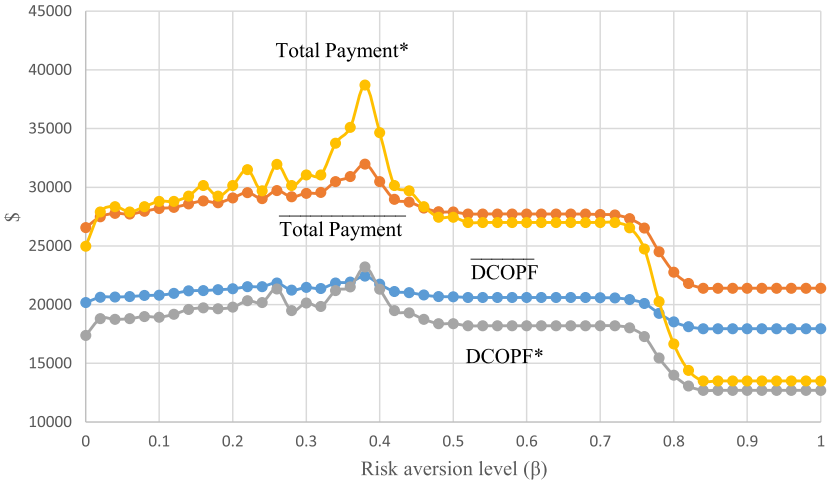



**Fig. 6.** DC-OPF optimal value and total payments to GenCos. (Presenting both end-of-simulation and simulation-average results.)

This analysis sheds light onto the effect of risk aversion level on GenCo bids and profits. Overall, while some level of risk aversion can be beneficial to GenCos' total profits, high levels of risk aversion is observed to degrade profits due to extreme price competition.

### 6.2. Differing risk aversion levels

Here, we analyze the effects of changes in the $\beta$ values of individual GenCos, focusing initially on GenCo-2. Fig. 7 presents the average profits and bid prices of each GenCo in a separate column, as a function of $\beta_2$ (in the *y* axis) and $\beta_3 = \beta_4$ (in the *x* axis) over the whole simulation. If $\beta_2$ increases, while keeping $\beta_3 = \beta_4$ constant, GenCo-2's profit decreases. This is expected as this GenCo now bids lower prices. Interestingly, GenCo-3's profit also decreases due to increased competition. GenCo-4's average profit, too, is generally reduced. The only exception with GenCo-4 occurs for very high $\beta_2$ values, in which GenCo-2 bids its minimum price 10 most frequently. In this case, GenCo-4 has a chance to make some profit only if $\beta_4$ is relatively low. Fig. 8 presents the end-of-simulation version of the same analysis.

Next, we investigate the effects of a simultaneous increase in $\beta_3 = \beta_4$, while keeping $\beta_2$ fixed, for example, at zero. When GenCos 3 and 4 become more risk-averse, they might be expected to reduce their bid prices, leading to a decrease in GenCo-2's profit. Our simulation results, however, suggest the opposite. As $\beta_3$ and $\beta_4$ increase, we observe GenCo-2 to increase its bid price, leading to an increase in its profit. The key to understanding this counter-intuitive result is GenCo-4's behavior, who simply bids its highest price alternative 45. For this price, GenCo-4 is assigned no dispatch. If GenCo-4 bids one of the lower prices, there is a slight chance that it will be assigned some dispatch and earn some profit. When this happens, however, the variability in profit also increases which is not desirable from a risk-aversion point of view.

Fig. 9 illustrates the average profit and bid price of each GenCo-*i* (in a separate column) over the simulation as a function of its own $\beta_i$ (in the *y* axis) and the other GenCos' $\beta$ values (in the *x* axis). The leftmost column is the same as that of Fig. 7. From the middle column, for example, we observe that rather than its own $\beta_3$, GenCo-3's profit depends mostly on the risk aversion levels of the other GenCos, particularly that of GenCo-2. GenCo-3 sticks to the advantageous bid price of 9 unless $\beta_3$ is very low. Given this $b_3$, GenCo-3's profit becomes a function of $b_2$, which decreases if $\beta_2$ increases. Note the emergence of $\beta = 0.38$ as a critical value again in this graph. GenCo-3 profits, in particular, are maximized when $\beta_2$ and $\beta_4$ are around 0.38. GenCo-4 makes a much lower profit compared to GenCos 2 and 3. GenCo-4's profit is maximized when $\beta_4$ is at intermediate values, while the other two GenCos' risk aversion levels are low, and consequently they do not engage in intense price competition.

Fig. 10 presents the end-of-simulation versions of these graphs. Compared to the simulation-average values, we observe GenCo-2 and GenCo-3's profits to be higher. GenCo-4's end-of-simulation profit, meanwhile, converges to zero independent of $\beta_4$.

### 6.3. Learning dynamics

Here, we drill further into the detailed workings of the learning model during the simulation. Fig. 11 presents how the three Gen-

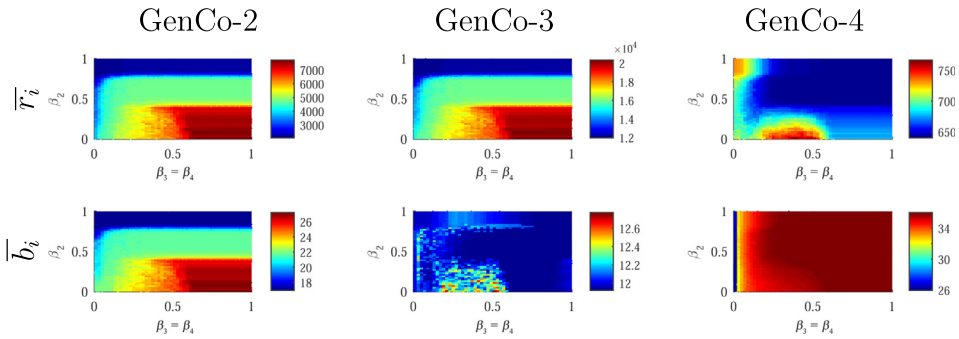



**Fig. 7.** Profits and bid prices as a function of $\beta_2$ vs. $\beta_3 = \beta_4$, simulation average.

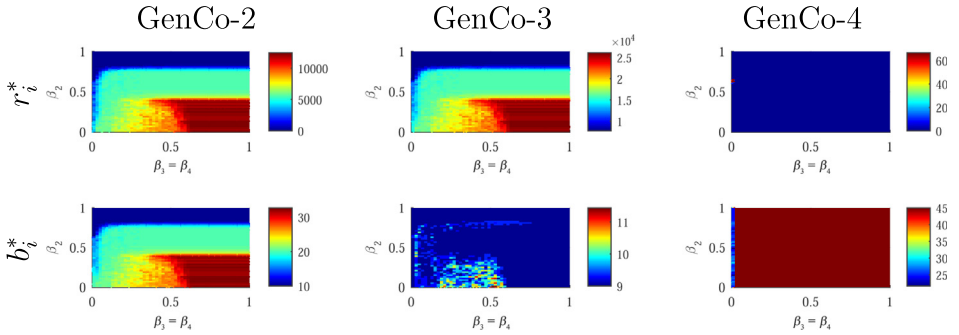



**Fig. 8.** Profits and bid prices as a function of $\beta_2$ vs. $\beta_3 = \beta_4$, end of simulation.

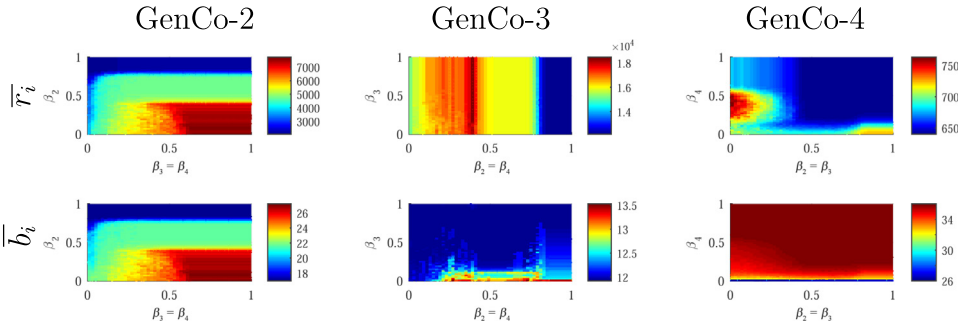



**Fig. 9.** Profits and bid prices as a function of risk aversion levels, simulation average.

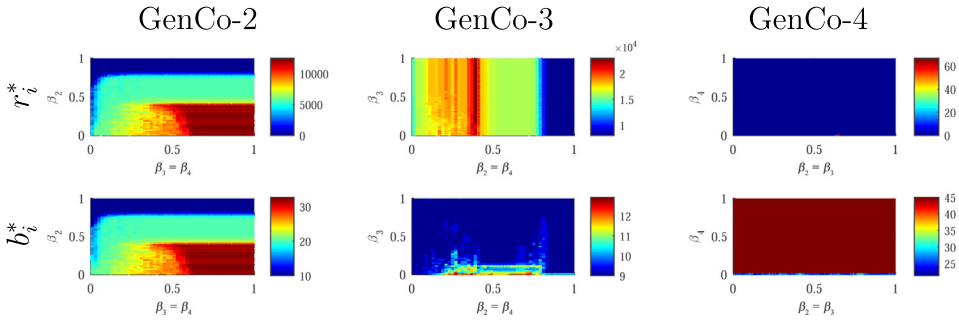



**Fig. 10.** Profits and bid prices as a function of risk aversion levels, end of simulation.

Cos' $Q^r$-values, hence the best identified bid prices, change over iterations for a given risk profile $(\beta_2, \beta_3, \beta_4)$. The graphs on left present the case of the risk profile $(0, 0, 0)$, corresponding to the bottom left corner of the relevant graph in Fig. 8. The graphs on right present the case of the risk profile $(0, 1, 1)$. Recall that our model ignores the effect of risk during the first half of the iterations; the risk model kicks in after iteration 1000.

When all GenCos are risk-neutral (Fig. 11(a)), we observe $b_2^*$ and $b_3^*$ to take some time to converge, due possibly to the tight competition between GenCos 2 and 3. Once the equilibrium between

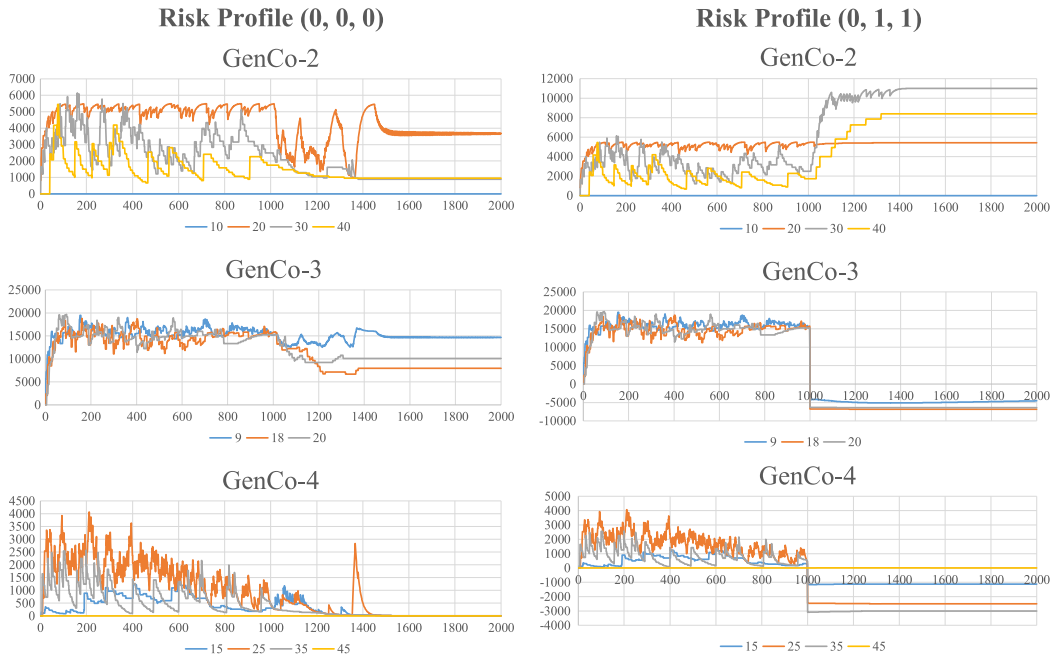



**Fig. 11.** $Q^r$-value evolutions. (a) Risk profile $(0,0,0)$. (b) Risk profile $(0,1,1)$.

these two GenCos with $b_2^* = 20$ and $b_3^* = 9$ is reached, GenCo-4's bids become irrelevant as this GenCo is driven out of the market.

When GenCo-2 is risk-neutral but GenCos 3 and 4 are extremely risk-averse (Fig. 11(b)), $b_2^*$ changes from 20 to 30 once risk aversion kicks in at iteration 1000. For GenCo-3, the price 9 arises as $b_3^*$. Bidding 9 brings in a decent profit to GenCo-3 while not having the profit variability disadvantage of the higher bid prices.

This discussion illustrates how the learning, risk aversion and competition components of our model interact with each other.

## 7. Effects of the Q-learning parameters

Here, we study the effects of the initial values, $\epsilon_{i0}$ and $\alpha_{i0}$, of the time-decaying $Q$-learning parameters, $\epsilon_{it}$ and $\alpha_{it}$, on GenCo profits. A comprehensive simulation study is conducted using the network structure of Case 3. In all simulation runs 2000 iterations are conducted.

We report the results from the perspective of one GenCo at a time (the *GenCo-i*), which is assumed to be risk neutral. For this GenCo, we consider $21 \times 21 = 441$ parameter combinations of $(\epsilon_{i0}, \alpha_{i0})$, in which each of the two parameters range between 0 and 1 with an increment size of 0.05.

For a given $(\epsilon_{i0}, \alpha_{i0})$ combination, we speak of different *scenarios* characterizing the parameters of the other two GenCos (GenCo-*k* where $k \neq i$) which are chosen from the following sets: $\epsilon_{k0} \in \{0.2, 0.4, 0.8\}, \alpha_{k0} \in \{0, 0.2, 0.8\}, \beta_{k0} \in \{0, 0.4, 0.8\}$. In each of the $3^3 \times 3^3 = 729$ considered scenarios, the same stream of random numbers are used and the results are averaged over 10 runs. GenCo-*i* is assumed to have no information about the parameters of the other GenCos; hence, it believes all scenarios to be equally likely. Consequently, for each $(\epsilon_{i0}, \alpha_{i0})$ combination of GenCo-*i*, the average $Q$-value $(\overline{Q_i})$ and the average cumulative profit $(\overline{CP_i})$ over all 729 scenarios are reported. All in all, this comprehensive simulation study required the DC-OPF problem to be solved 19,289,340,000 times (3 GenCo-*i* $\times$ 441 combinations $\times$ 729 sce-

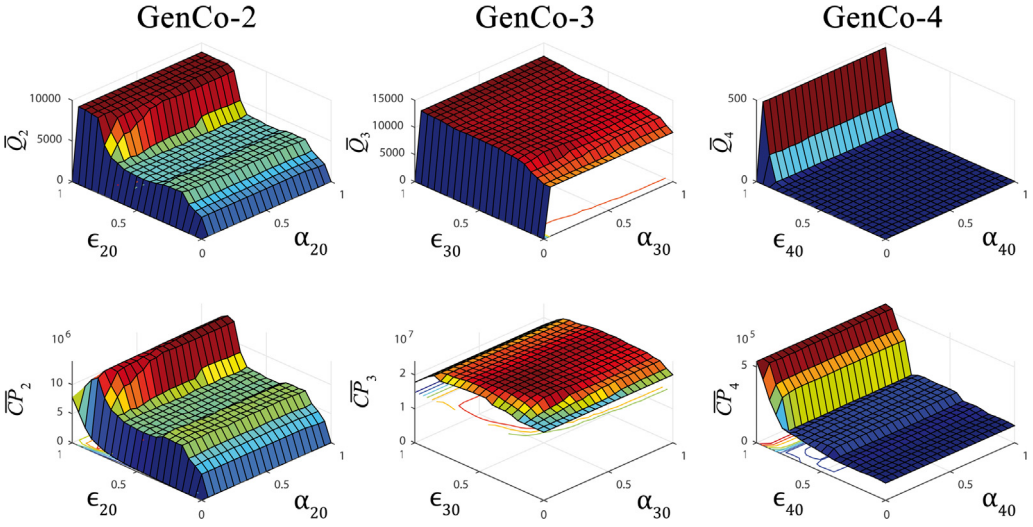



**Fig. 12.** $\overline{Q_i}$ (first row) and $\overline{CP_i}$ (second row) for different $(\alpha_{i0}, \epsilon_{i0})$ combinations.

narios × 10 runs × 2000 iterations). The study took around 2000 h on an Intel Core i7 @ 3.2 GHz computer with 24 GB RAM.

Fig. 12 presents the results (averaged over 729 scenarios) for each GenCo-*i* in a separate column. Graphs in the first row illustrate GenCo-*i*'s expected profit, that is, the *Q*-value of the best identified bid at the end of the simulation, whereas those in the second row illustrate the cumulative profit (CP) throughout the simulation. From the figure, we observe $\alpha_{i0}$ not to have a major impact on profit results unless its value is very low. Thus, the profits are robust to the initial value of the recency rate as long as some updating of *Q*-values occur. The exploration parameter $\epsilon_{i0}$, on the other hand, is seen to have a significant impact on profits. The direction of this impact, however, is ambiguous. For GenCo-2, high $\epsilon_{20}$ lead to better profits. For GenCo-3, this is true for the end-of-simulation profit, yet the cumulative profit first increases then decreases with $\epsilon_{30}$. For GenCo-4, profit is uniformly increasing in $\epsilon_{40}$. Recall that GenCo-4 is at a disadvantageous position compared to the other GenCos. As suggested in Fig. 12, this GenCo can maximize its expected profit by acting as randomly as possible (corresponding to $\epsilon_{40} = 1$), thereby disrupting the learning of the two other GenCos.

Next, we compare the end-of-simulation and cumulative profit values. For GenCos 2 and 4, $(\epsilon_{i0}, \alpha_{i0})$ combinations that yield the highest expected profit at the end of the simulation (the first row graphs) are also observed to provide the highest cumulative profit (the second row graphs). For GenCo-3, on the other hand, we observe significant differences. This GenCo can identify better profit opportunities at the end of the simulation by exploring excessively, however, this comes at the cost of achieving a lower cumulative profit. Overall, the only parameter that has a major profit impact at the end of the simulation turn out to be $\epsilon_{20}$. High values of this parameter is observed to increase the expected profit of GenCo-2 significantly.

## 8. Discussion

The academic literature on electricity markets is somewhat overly concerned with theoretical issues such as convergence to Nash equilibria [61]. Market participants, meanwhile, need studies that address the profits and risks associated with realistic GenCo bidding strategies. In fact, one of the research suggestions in Dahlgren et al. [31]'s survey on risk assessment in energy trading is "risk assessment would be more accurate if the bidding behaviors of market players can be modeled". This is what we do in the present study by considering the effects of two behavioral factors on GenCo bidding and market evolution: learning from experience, and risk aversion. Our findings imply that one should be cautious in using static models to investigate dynamic markets such as the day-ahead electricity market. This is because these models fail to capture the dynamics of the interaction between competing GenCos that learn from experience.

We present the first ABMS study to analyze the joint effects of learning and risk aversion, leading to interesting observations. In particular, different from the literature, we obtain non-monotonous results concerning the effect of risk aversion on profits. As Liu and Wu [57] mention, most studies in GenCo risk management literature (e.g., [46,62,63]) neglect the market dynamics by considering known (or fixed) probabilistic distributions for the price of electricity, demand, or rivals' behavior. Consequently, these models generally find a monotonous decrease in profit and the taken risk as GenCos become more risk averse. In practice, such results may not hold true since the interaction of learning GenCos can affect the distribution of the aforementioned factors through time [50,51].

Our results have important regulatory and managerial implications. Most importantly, simulation models such as our model can be used in the development of *testbeds* that are tailored to the learning behavior and risk aversion levels of GenCos in a particular electricity market. For instance, a large and established GenCo with a strong financial status would be modeled as less risk averse than a small GenCo. Risk aversion levels can be modeled as time-dependent, to capture the changes in risk attitude due to changing economic and financial conditions. Likewise, the learning behavior of GenCos would reflect the overall firm strategy, experience in the market and capabilities of the firm's human resources. For example, a conservative GenCo would be modeled with a smaller $\epsilon$ parameter value compared to a GenCo that is more open to trying alternative bid prices.

ISOs may use the aforementioned testbeds to develop market rules that manipulate GenCo behavior in directions that provide higher social welfare. Such testbeds would allow an ISO to study the likely impacts of market rule changes prior to costly real market implementation. These studies would be particularly important in the prevention of *tacit collusion* among GenCos that can arise from GenCo learning [29,64,65]. GenCos, too, would benefit from testbeds, in formulating bidding policies that consider their own as well as their competitors' bidding behavior.

## 9. Conclusions

This study analyzes how learning dynamics and risk aversion shape GenCo bidding behavior in a competitive electricity market, using an agent-based simulation model. GenCos are modeled as agents that bid prices repeatedly for each hour of the day-ahead market. Learning is modeled through a modified *Q*-learning algorithm, and risk aversion is captured as aversion to variability in profits. Given GenCos' bids, to determine locational marginal prices and GenCo power dispatches, the ISO solves a DC-OPF problem that considers the physical network characteristics.

First, considering risk-neutral GenCos, the mechanics of our learning algorithm is illustrated on two simple case studies. The simulation runs achieve convergence thanks to time-decaying *Q*-learning parameters. In the case of a unique Nash equilibrium, the simulation easily converges to the equilibrium. In the presence of multiple equilibria, however, simulation runs converge to either one of the Nash equilibria, or a state that provides identical profits to a Nash equilibrium. Thus, the individual learning of GenCos is observed to drive the market into a reasonable outcome.

When the model is extended to consider risk-averse GenCos, the results show that some level of risk aversion may indeed be beneficial for GenCo total profits compared to the risk-neutral case. On the other hand, high levels of risk aversion are observed to intensify price competition and degrade profits. The findings illustrate how altering the risk aversion level of even one GenCo can trigger changes in the bidding behavior and profit levels of all GenCos due to learning and market interaction.

This study can be extended in numerous directions. First and foremost, one could use other measures of risk such as CVaR. A different learning model might be used, or the *Q*-learning algorithm employed in this study may be extended to achieve better performance. For instance, a GenCo may improve its profit by considering changes in the *Q*-values in addition to the *Q*-values themselves. Definitely, the results may depend on the specific risk aversion or learning method being used, which calls for follow-up studies.

## References

[1] David AK, Wen F. Strategic bidding in competitive electricity markets: a literature survey. Power engineering society summer meeting 2000, vol. 4. IEEE; 2000. p. 2168–73.
[2] Hobbs BF, Metzler CB, Pang JS. Strategic gaming analysis for electric power systems: an MPEC approach. IEEE Trans Power Syst 2000;15(2):638–45.

[3] Conejo AJ, Carrión M, Morales JM. Decision making under uncertainty in electricity markets, vol. 1. Springer; 2010.
[4] Li G, Shi J, Qu X. Modeling methods for GenCo bidding strategy optimization in the liberalized electricity spot market – a state-of-the-art review. Energy 2011;36(8):4686–700.
[5] Krause T, Andersson G, Ernst D, Vdovina-Beck E, Cherkaoui R, Germond A. Nash equilibria and reinforcement learning for active decision maker modelling in power markets. In: Proceedings of the 6th IAEE European conference: modelling in energy economics and policy. p. 1–6.
[6] Krause T, Beck EV, Cherkaoui R, Germond A, Andersson G, Ernst D. A comparison of Nash equilibria analysis and agent-based modelling for power markets. Int J Electr Power Energy Syst 2006;28(9):599–607.
[7] Aliabadi DE, Kaya M, Şahin G. An agent-based simulation of power generation company behavior in electricity markets under different market-clearing mechanisms. Energy Policy 2017;100:191–205.
[8] Ventosa M, Baıllo A, Ramos A, Rivier M. Electricity market modeling trends. Energy Policy 2005;33(7):897–913.
[9] Ruiz C, Conejo A, García-Bertrand R. Some analytical results pertaining to Cournot models for short-term electricity markets. Electr Power Syst Res 2008;78(10):1672–8.
[10] Bunn DW, Oliveira FS. Evaluating individual market power in electricity markets via agent-based simulation. Ann Oper Res 2003;121(1–4):57–77.
[11] Li T, Shahidehpour M. Strategic bidding of transmission-constrained GenCos with incomplete information. IEEE Trans Power Syst 2005;20(1):437–47.
[12] Day CJ, Hobbs BF, Pang JS. Oligopolistic competition in power networks: a conjectured supply function approach. IEEE Trans Power Syst 2002;17(3):597–607.
[13] Díaz CA, Villar J, Campos FA, Reneses J. Electricity market equilibrium based on conjectural variations. Electr Power Syst Res 2010;80(12):1572–9.
[14] Ruiz C, Conejo AJ, Arcos R. Some analytical results on conjectural variation models for short-term electricity markets. IET Gener Transm Distrib 2010;4(2):257–67.
[15] Weidlich A. Engineering interrelated electricity markets: an agent-based computational approach. Springer Science & Business Media; 2008.
[16] Ruiz C, Kazempour SJ, Conejo AJ. Equilibria in futures and spot electricity markets. Electr Power Syst Res 2012;84(1):1–9.
[17] Kardakos EG, Simoglou CK, Bakirtzis AG. Optimal bidding strategy in transmission-constrained electricity markets. Electr Power Syst Res 2014;109:141–9.
[18] Weidlich A, Veit D. A critical survey of agent-based wholesale electricity market models. Energy Econ 2008;30(4):1728–59.
[19] Bunn DW, Oliveira FS. Agent-based simulation-an application to the new electricity trading arrangements of England and Wales. IEEE Trans Evol Comput 2001;5(5):493–503.
[20] Veit DJ, Weidlich A, Krafft JA. An agent-based analysis of the German electricity market with transmission capacity constraints. Energy Policy 2009;37(10):4132–44.
[21] Li G, Shi J. Agent-based modeling for trading wind power with uncertainty in the day-ahead wholesale electricity markets of single-sided auctions. Appl Energy 2012;99:13–22.
[22] Erev I, Roth AE. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. Am Econ Rev 1998;88(4):848–81.
[23] Watkins CJ, Dayan P. Q-learning. Mach Learn 1992;8(3–4):279–92.
[24] Sutton RS. Learning to predict by the methods of temporal differences. Mach Learn 1988;3(1):9–44.
[25] Nicolaisen J, Petrov V, Tesfatsion L. Market power and efficiency in a computational electricity market with discriminatory double-auction pricing. IEEE Trans Evol Comput 2001;5(5):504–23.
[26] Li H, Tesfatsion L. Co-learning patterns as emergent market phenomena: an electricity market illustration. J Econ Behav Org 2012;82(2):395–419.
[27] Sharbafi MA, Azidehak A, Hoshyari M, Babarsad OB, Aliabadi DE, Zareian A, et al. MRL extended team description 2011. In: Proceedings of the 15th international RoboCup symposium, Istanbul, Turkey. p. 1–29.
[28] Wang J. Conjectural variation-based bidding strategies with Q-learning in electricity markets. In: 42nd Hawaii international conference on system sciences, 2009, HICSS'09. IEEE; 2009. p. 1–10.
[29] Yu NP, Liu CC, Price J. Evaluation of market rules using a multi-agent system method. IEEE Trans Power Syst 2010;25(1):470–9.
[30] Guo M, Liu Y, Malec J. A new q-learning algorithm based on the metropolis criterion. IEEE Trans Syst Man Cybernet Part B (Cybernet) 2004;34(5):2140–3.
[31] Dahlgren R, Liu CC, Lawarree J. Risk assessment in energy trading. IEEE Trans Power Syst 2003;18(2):503–11.
[32] Bathurst GN, Weatherill J, Strbac G. Trading wind generation in short term energy markets. IEEE Trans Power Syst 2002;17(3):782–9.
[33] Ni E, Luh PB, Rourke S. Optimal integrated generation bidding and scheduling with risk management under a deregulated power market. IEEE Trans Power Syst 2004;19(1):600–9.
[34] Conejo AJ, Garcia-Bertrand R, Carrion M, Caballero Á, de Andres A. Optimal involvement in futures markets of a power producer. IEEE Trans Power Syst 2008;23(2):703–11.
[35] Morales JM, Conejo AJ, Pérez-Ruiz J. Short-term trading for a wind power producer. IEEE Trans Power Syst 2010;25(1):554–64.
[36] Morales JM, Conejo AJ, Madsen H, Pinson P, Zugno M. Trading stochastic production in electricity pools. In: Integrating renewables in electricity markets. Springer; 2014. p. 205–42.
[37] Zheng QP, Wang J, Liu AL. Stochastic optimization for unit commitment – a review. IEEE Trans Power Syst 2015;30(4):1913–24.
[38] Conejo AJ, Nogales FJ, Arroyo JM, García-Bertrand R. Risk-constrained self-scheduling of a thermal power producer. IEEE Trans Power Syst 2004;19(3):1569–74.
[39] García-González J, Parrilla E, Mateo A. Risk-averse profit-based optimal scheduling of a hydro-chain in the day-ahead electricity market. Eur J Oper Res 2007;181(3):1354–69.
[40] Dicorato M, Forte G, Trovato M, Caruso E. Risk-constrained profit maximization in day-ahead electricity market. IEEE Trans Power Syst 2009;24(3):1107–14.
[41] Ghadikolaei HM, Ahmadi A, Aghaei J, Najafi M. Risk constrained self-scheduling of hydro/wind units for short term electricity markets considering intermittency and uncertainty. Renew Sustain Energy Rev 2012;16(7):4734–43.
[42] Jiang R, Wang J, Guan Y. Robust unit commitment with wind power and pumped storage hydro. IEEE Trans Power Syst 2012;27(2):800–10.
[43] Nojavan S, Zare K. Risk-based optimal bidding strategy of generation company in day-ahead electricity market using information gap decision theory. Int J Electr Power Energy Syst 2013;48:83–92.
[44] Chen J, Zhuang Y, Li Y, Wang P, Zhao Y, Zhang C. Risk-aware short term hydro-wind-thermal scheduling using a probability interval optimization model. Appl Energy 2017;189:534–54.
[45] Batlle C, Otero-Novas I, Alba J, Meseguer C, Barquín J. A model based in numerical simulation techniques as a tool for decision-making and risk management in a wholesale electricity market. Part I: General structure and scenario generators. In: PMAPS00 Conference. p. 1–7.
[46] Gountis VP, Bakirtzis AG. Bidding strategies for electricity producers in a competitive electricity marketplace. IEEE Trans Power Syst 2004;19(1):356–65.
[47] Caruso E, Dicorato M, Minoia A, Trovato M. Supplier risk analysis in the day-ahead electricity market. IEE Proc Gener Transm Distrib 2006;153(3):335–42.
[48] Cabero J, Ventosa MJ, Cerisola S, Baillo A. Modeling risk management in oligopolistic electricity markets: a Benders decomposition approach. IEEE Trans Power Syst 2010;25(1):263–71.
[49] Pousinho HM, Contreras J, Bakirtzis AG, Catalão JP. Risk-constrained scheduling and offering strategies of a price-maker hydro producer under uncertainty. IEEE Trans Power Syst 2013;28(2):1879–87.
[50] Chin D, Siddiqui A. Capacity expansion and forward contracting in a duopolistic power sector. Comput Manage Sci 2014;11(1–2):57–86.
[51] Egging R, Pichler A, Kalvø ØI, Walle-Hansen TM. Risk aversion in imperfect natural gas markets. Eur J Oper Res 2017;259(1):367–83.
[52] Fleten SE, Wallace SW, Ziemba WT. Hedging electricity portfolios via stochastic programming. In: Decision making under uncertainty. Springer; 2002. p. 71–93.
[53] Vehviläinen I, Keppo J. Managing electricity market price risk. Eur J Oper Res 2003;145(1):136–47.
[54] Sahin C, Shahidehpour M, Erkmen I. Generation risk assessment in volatile conditions with wind, hydro, and natural gas units. Appl Energy 2012;96:4–11.
[55] Gielis F. Potential effects of risk aversion on technology choices and security of supply: researched with an agent-based model of a liberalised electricity market Ph.D. thesis. TU Delft: Delft University of Technology; 2016.
[56] Di Lorenzo G, Pilidis P, Witton J, Probert D. Monte-carlo simulation of investment integrity and value for power-plants with carbon-capture. Appl Energy 2012;98:467–78.
[57] Liu Y, Wu F. Risk management of generators' strategic bidding in dynamic oligopolistic electricity market using optimal control. IET Gener Transm Distrib 2007;1(3):388–98.
[58] Rahimiyan M, Mashhadi HR. An adaptive-learning algorithm developed for agent-based computational modeling of electricity market. IEEE Trans Syst Man Cybernet Part C (Appl Rev) 2010;40(5):547–56.
[59] Sun J, Tesfatsion L, et al. DC optimal power flow formulation and solution using QuadProgJ. In: Proceedings, IEEE power and energy society general meeting, Tampa, Florida. p. 1–62.
[60] Sutton RS, Barto AG. Reinforcement learning: an introduction, vol. 1. Cambridge: MIT Press; 1998.
[61] Shoham Y, Powers R, Grenager T. Multi-agent reinforcement learning: a critical survey. Tech rep; 2003.
[62] Boonchuay C, Ongsakul W. Optimal risky bidding strategy for a generating company by self-organising hierarchical particle swarm optimisation. Energy Convers Manage 2011;52(2):1047–53.
[63] Ma X, Wen F, Ni Y, Liu J. Towards the development of risk-constrained optimal bidding strategies for generation companies in electricity markets. Electr Power Syst Res 2005;73(3):305–12.
[64] Liu AL, Hobbs BF. Tacit collusion games in pool-based electricity markets under transmission constraints. Math Program 2013;140(2):351–79.
[65] Aliabadi DE, Kaya M, Şahin G. Determining collusion opportunities in deregulated electricity markets. Electr Power Syst Res 2016;141:432–41.