This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

An Agent-Based Framework for Policy Simulation: Modeling Heterogeneous Behaviors With Modified Sigmoid Function and Evolutionary Training

Songmin Yu

***Abstract*— This article proposes an agent-based policy simulation framework that can be applied to the cases satisfying: 1) the agents try to maximize some intertemporal preference and 2) the impacts of different factors on agents' behavioral tendency are monotonic. By combining the simulation and optimization methods, this framework balances the flexibility and validity of agent-based models (ABMs): the sigmoid function is modified and used to model agents' decision-making rules, and the evolutionary training method is used to calibrate agents' behavioral parameters. Based on an example for the emission trading scheme, the application of the framework is presented and evaluated in detail.**

***Index Terms*— Agent-based model (ABM), evolutionary training, flexibility and validity, policy simulation framework, sigmoid function.**

I. Introduction

AGENT-BASED models (ABMs) characterize socioeconomic systems as dynamic interactions among agents from a bottom-up perspective [1] and can introduce more complex interaction mechanisms in policy simulation, for example, herding effect in the crowd behavior [2], [3], continuous double auction mechanism in the financial markets [4], impact of network in the social choice [5], and society and collaboration [6]. Furthermore, ABMs can consider the attribute and behavior heterogeneity among agents [7], [8], as well as the bounded rationality and adaptation in agents' decision-making processes [9].

However, these modeling flexibilities come with two "major costs." First is the lack of a general framework compared with the neoclassical economic models, especially for modeling agents' decision-making processes, which can range from "zero-intelligent agents who act randomly [10]" to "sophisticated agents who can forecast, optimize, and adapt their strategies [9]." This wide range hinders the communication in the ABM community, reduces the comparability between different models, and also leads to the second cost: the difficulties in ensuring the model validity.

For comparison, neoclassical economic models follow a general theoretical setup (i.e., agents' utility optimization and market equilibrium) and locate their validity first in rigorous mathematical derivation and then in their consistency with empirical data. Perfect rationality might not be a perfect assumption for agents [11], but it disciplines the modeling of agents' behaviors and is useful and testable. On the other hand, ABMs model and simulate agents' decision-making processes and interactions following the bounded rationality. Ideally, the validity for modeling the behaviors of agents is built on: 1) validated behavioral theory [12] or 2) empirical calibration based on human experiment [13], [14], survey data [15], and so on. However, this theoretical support or empirical evidence is not always available.

In this article, the author tries to respond to both "major costs" of ABMs by proposing an agent-based framework for policy simulation, which can be applied to the cases satisfying the following two conditions: 1) the agents try to maximize some intertemporal preference [16]—for example, accumulated profit through the simulation periods—by making repetitive decisions in each period and 2) the impacts of different factors on agents' behavioral tendency are monotonic, for example, lower cost and higher social preference will increase a consumer's tendency to buy a specific car.

The framework includes two parts as follows.

1) First is a flexible and interpretable function form for modeling agents' decision-making rules, i.e., their policy functions for the repetitive decisions. By modifying the sigmoid function $f(x) = 1/(1+e^{-x})$ for different situations, the function form can integrate multiple impacts on agents' decisions flexibly (see Section III-A).
2) Second is an evolutionary training method for calibrating the "behavioral parameters" in agents' policy functions. Specifically, a "training stage" is introduced to initialize the "behavioral parameters" of the agents, and then, in the "simulation stage," the impact of policies is simulated and evaluated. Without relying on the empirical data, the agents learn from their past experience and update the behavioral parameters for higher intertemporal utility (see Section III-B).

In summary, this framework tries to balance the modeling flexibility and validity by combining the simulation and optimization methods. The flexibility is kept by constructing agents' decision-making rules based on the sigmoid function. The validity is improved (or partially guaranteed) by training the agents beforehand—they are still with bounded rationality

Manuscript received 21 March 2022; revised 26 May 2022, 19 July 2022, and 26 July 2022; accepted 3 August 2022.

The author is with the Fraunhofer Institute for Systems and Innovation Research, 76139 Karlsruhe, Germany (e-mail: songmin.yu@isi.fraunhofer.de).

Digital Object Identifier 10.1109/TCSS.2022.3196737

in the simulation, but to a certain extent, are making reasonable decisions. Besides, the quality and impact of the training stage can be evaluated by: 1) observing the convergence in the training stage, including agents' behavioral parameters at the micro level and the simulation results at the macro level and 2) comparing the simulation results with and without the training stage.

The rest of this article is organized as follows. Section II reviews the studies on ABM design and calibration and discusses the tradeoff between flexibility and validity in using ABMs for policy simulation. As a balancing solution, the proposed framework is introduced in Section III. Then, Section IV provides an application example of the framework based on my previous study on the carbon emission trading scheme (ETS) [8], with a special focus on the implementation and evaluation of the evolutionary training in detail. Finally, this article is concluded in Section V.

II. Literature Review

All models are simplifications. We can build wind tunnels to simulate the flying environment for designing aircraft based on aerodynamics. However, we have to simplify more when modeling the socioeconomic systems: what we do not satisfactorily know are not only the parameters but also the mechanisms. In this regard, model choice is fundamental perspective choice: it depends on which mechanism the modeler wants to emphasize.

This fact does not sound scientific but implies two fundamental criteria for model selection. First is flexibility, meaning that the model should be handy enough to capture the important mechanism for explaining the phenomenon in focus. Second is validity, meaning that the model is acceptable for its intended use because it meets specified performance requirements [17]. When developing models for policy simulation, the tension is always to improve the flexibility without losing too much validity and vice versa. Compared with the mainstream neoclassical economic models, ABMs have been growing as an alternative and supplement by following a new paradigm to balance the flexibility and validity.

1) Regarding the flexibility, ABMs model agents' behaviors with bounded rationality as: 1) a set of if–else rules; 2) a response function (i.e., "fast and frugal heuristics") of their states and the signals from the environment [4], [18]; or 3) myopic optimization based on imperfect information and limited computation capacity [9]. Thus, without the pressure to solve agents' rational decisions in a dynamic setting, ABMs can flexibly integrate: 1) details of the environment, e.g., agents' interaction mechanism, coevolution of the environment and agents, spatial or network structure, and so on; 2) attribute or behavior heterogeneity among agents; and 3) agents' learning and adaptation behaviors.
2) Regarding the validity, ABMs model agents' behaviors based on: 1) validated theories of boundedly rational decision-making process, e.g., the behavioral theory of the firm [12], [19], [20], [21], or 2) existing framework calibrated with empirical data from human experiments, e.g., discrete choice experiment [15], [22] and simple forecasting heuristics [13], [14]. Besides, the evolutionary training method can also be used to enhance model validity [8].

TABLE I

Aspects for Designing an Agent-Based Market Model

	Mechanism	Parameter
Environment	price formation mechanism	total amount of product, tax, subsidy, etc.
Agent	decision-making with bounded rationality	endowment, preference, risk attitude, etc.

In the following, studies on ABM design (Section II-A) and calibration (Section II-B) are reviewed to discuss the tradeoff frontier between flexibility and validity in using ABMs for policy simulation and, further, to identify the cases for which we can move the frontier forward by using the framework proposed in this study: modeling heterogeneous behavior with modified sigmoid function and evolutionary training.

A. Flexibility: ABM Design

Designing an ABM includes two parts. First is the environment, meaning the resources and constraints in the system, the interaction mechanism among agents, and agents' impact on the environment (i.e., coevolution between the environment and agents). Second is the agents, including their attributes and decision-making rules. Taking an agent-based market model as an example, the designing aspects are summarized in Table I.

Concerning the environment design, more price formation mechanisms other than "market clearing" can be modeled by the ABMs to capture specific price dynamics [16], for example, market maker [23], floor trading [24], and continuous double auction [4], [10]. In other cases, agents can also interact by exchanging information [25], imitating behavior [2], [26], [27], learning from each other [28], [29], and so on. Introducing the interaction mechanisms in reality to the environment can be one major reason why ABM is used, as interaction is the fundamental reason that makes a system "complex," i.e., with emergent properties.

Concerning the agents design, on one hand, ABMs take the bottom-up perspective and can flexibly consider the heterogeneity among agents, concerning their attributes (e.g., endowment and preference) or behaviors (e.g., the "fundamentalists" or "chartists" in financial ABMs [7]). By considering the heterogeneity among agents, two types of questions can be answered: 1) how can the existence and degree of heterogeneity influence the system dynamics? and 2) what are the distributional impacts of a specific policy design on the heterogeneous agents in the system?

On the other hand, ABMs can model agents' bounded rationality in their decision-making processes, which provide interfaces to psychological and behavioral studies. Fig. 1 shows a full coverage of the modules to construct an agent in the ABMs, including the information flow. Each of the three modules—cognitive architecture, decision formation, and

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

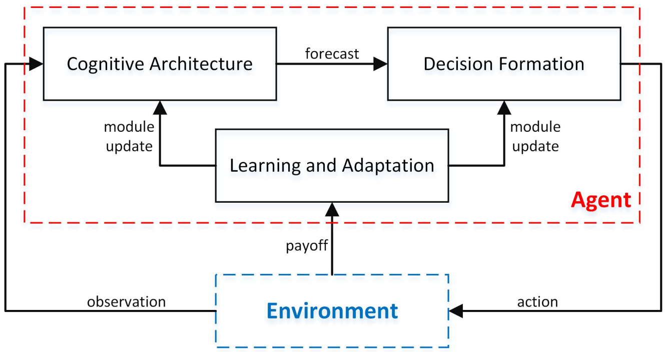

Fig. 1. Design of agents in the ABMs.

learning and adaption—has different implementations in different ABMs as long as they can fit together. Besides, for different systems and research questions, it is not always necessary to include all the modules. From simple to sophisticated, four modes for designing the agents are summarized as follows.

First is zero-intelligence mode, meaning that the agents have no "cognitive architecture" or "learning and adaption" modules and only have almost empty "decision formation" modules that are as simple as "random behavior." Gode and Sunder [10] designed "zero-intelligence" agents that randomly draw bidding or asking prices within the budget constraints and trade with each other following the continuous double auction mechanism to explore the allocative efficiency of the market mechanism.[1]

Second is responsive mode, meaning that the agents observe the environment and react directly based on their "decision formation" modules without forming any forecast based on the "cognitive architecture." Besides, the agents do not have "learning and adaption" modules, either. The "decision formation" module can be implemented in different ways: 1) a set of "if–else" rules [12]; 2) "fast and frugal heuristics" [18] represented by behavioral functions [8] or artificial neural networks [30]; and 3) myopic optimization.

Third is deliberative mode, meaning that the agents first form the "forecast of the future" based on their "cognitive architecture" modules, and then, the forecasts are taken as input for their "decision formation" modules to finally form their decisions. Similar to the "decision formation" module, there are also different implementations for the "cognitive architecture" module [31], [32].

Fourth is introspective mode, meaning that the agents can update their "cognitive architecture" and/or "decision formation" modules based on their "learning and adaption" modules. Conceptually, learning and adaptation are parts of behavior rules, which concern how the behavior rules are updated. The design of "learning and adaption" module highly depends on the design of the module(s) it aims to update. Brenner [33] distinguished between "nonconscious learning," "routine-based learning," and "belief learning" in the ABMs.[2] Furthermore, another distinguishing criterion is the population level of learning [29]: 1) individual learning: each agent learns only based on its own experience [9] and 2) social learning: agents also take experiences of other agents (behaviors and payoffs) into consideration during the learning process [36], [37]. Table II summarizes different methods for implementing the "learning and adaption" module in ABMs.[3]

B. Validity: ABM Calibration

Manson [46] summarized two steps in developing an ABM for a target system.

1) Distill the system into a "conceptual model" by identifying the relevant "mechanisms" of the environment or agents.
2) As shown in Table I, instantiate the conceptual model in a "software model" (i.e., the ABM) by: 1) implementing the mechanisms with specific function forms based on existing theories or assumptions and 2) calibrating the parameters based on empirical data of the target system.

For a developed ABM, "validity" refers to the overall quality of the model. "Validation" means to assess the model validity according to specified performance requirements that the model needs to satisfy for its intended use [17]. To improve the validity, existing studies mainly focus on parameter calibration, including two categories of methods.

The first category is direct calibration, which is to assign values for model parameters directly based on empirical data, including direct observation [47], analytical methods [48], survey [15], and human experiment [13], [14]. However, due to the limitations of data availability (especially at the agent level) and complexities of the model, very few ABMs can be calibrated (only) based on the direct calibration methods [49]. The second category of methods is simulation-based calibration, which can be further categorized as: 1) simulated

[1]By comparing the "allocative efficiency" of the trading among "zero-intelligent" agents and the trading among human traders, the authors found that the difference between the two cases is not significant (both close to 100%). Then, the authors concluded that "allocative efficiency" of a continuous double auction derives largely from its structure, independent of traders' motivation, intelligence, or learning.

[2]"Nonconscious learning" corresponds to the situation when learning aims to update the "decision formation" module that is designed as follows: the agent faces N possible actions, and the decision-making rule is modeled as a choice probability distribution defined on the actions. Then, the agents update their choice probability distribution upon the following logic: if an action leads to a negative outcome, it will be avoided in the future; while if an action leads to a positive outcome, it will reoccur [34]. "Routine-based learning" corresponds to the situation when the "learning and adaption" module aims to update the "decision formation" module: a direct connection is established from agents' past experiences and observations to their current behaviors, according to some fundamental principles of learning concluded from experiment results or *ad hoc* reasoning [35]. "Belief learning" corresponds to the situation when the "learning and adaption" module aims to update the "cognitive architecture" module [9]. "Belief learning" relates to real learning process most, which is also referred to as "cognitive learning" in the psychology literature, while routine-based learning is only designed to represent certain features of learning processes approximately.

[3]The meanings of the abbreviations are as follows: NL, RL, and BL refer to "nonconscious learning," "routine-based learning," and "belief learning"; I/S refers to "individual learning/social learning"; SPGA refers to "single-population genetic algorithm"; SPGP refers to "single-population genetic programming"; MPGA refers to "multiple-population genetic algorithm"; and MPGP refers to "multiple-population genetic programming."

TABLE II
METHODS FOR IMPLEMENTING THE "LEARNING AND ADAPTATION" MODULE IN ABMS

Methods	NL	RL	BL	I / S	Introduction
Reinforcement learning	✓		✓	I	Agents update the choice probability distribution defined on actions (or beliefs) based on an updating rule and payoff (or forecast accuracy) [34].
Strategy switching	✓		✓	I	Agents probabilistically switch among actions (or beliefs) according to the payoff (or forecast accuracy) [38].
Experimentation learning		✓		I	Agents update their behavior based on the trial-and-error principle [39].
Melioration learning		✓		I	Agents update the probability distribution defined on their actions based on their average payoff in the past [40].
Imitation		✓		S	Agents observe the actions and payoffs of others, then update their behavioral rules accordingly [35].
Satisficing		✓		I	Each agent has an aspiration level. When the the actual payoff is above it, they are satisfied. Otherwise they will probabilistically alter their strategies [41].
Replicator dynamics		✓		S	The actions that lead to payoff higher than average will occur more frequently [42].
SPGA		✓	✓	S	The behavioral rule (or belief) of each agent is represented by a chromosome, and learning is implemented based on the Genetic Algorithm within the society [37].
SPGP		✓	✓	S	The behavioral rule (or belief) of each agent is represented by a tree, and learning is implemented based on the Genetic Programming within the society [37].
MPGA		✓	✓	I	The behavioral rule (or belief) of each agent is represented by a society of chromosomes, and learning is implemented by running the Genetic Algorithm inside each agent [8].
MPGP		✓	✓	I	The behavioral rule (or belief) of each agent is represented by a society of trees, and learning is implemented by running the Genetic Programming inside each agent [37].
Fictitious play			✓	I	Agents collect the past experiences, and update their beliefs assuming that the likelihood of these events and causal relations is given by a stationary probability distribution [43].
Bayesian learning			✓	I	Agents' beliefs of the environment are represented by a set of complete and complementary hypotheses. Agents update their probability according to the Bayesian law [44].
Least-square learning			✓	I	The belief of each agent is represented by a function with a set of parameters. Then, agents fit the parameters such that the sum of the squares of the differences between the predicted and the observed values becomes minimal [45].
Neural network	✓		✓	I	The decision-making rule or belief of each agent is represented by a neural network, and agents update the parameters of the network based on the payoff or forecast accuracy [36].
Stochastic belief learning			✓	I	Agents hold a small number of beliefs for forecasting, which is a subset of all possible beliefs. Then, this subset is updated according to incoming information [33].

minimum distance (SMD) methods[4]; 2) likelihood-based methods [55], [56]; 3) Bayesian approaches [57]; and 4) abduction analysis[5] [58]. Platt [49] compared a number of methods to determine the respective strengths and weaknesses of each approach. Fagiolo *et al.* [59] provided a critical review of the existing validation techniques based on a simple theoretical framework.

[4]Generally, the idea of SMD can be summarized as three steps: 1) run the model with specific parameter combinations (randomly drawn for start) and produce simulated data; 2) measure the "distance" between the simulated and observed data; and 3) use the measured distance to drive searching the parameter space, recursively run the first two steps, and ultimately find the parameter combination that minimize the distance. Regarding the measurement of distance, several criteria or methods have been proposed, including the method of simulated moments (MSM) [50], indirect inference (II) [51], sum of appropriately weighted squared differences [52], Markov information criterion (MIC) [53], and generalized subtracted L-divergence (GSL-div) [54].

[5]Compared with the first three approaches, "abduction analysis" does not insist on finding specific "values" for the parameters but tries to reduce the "parameter space" as far as the empirical data allows. First, a set of parameter vectors are randomly sampled from the initial "parameter space," and each of them corresponds to a "model specification." Then, a statistical test is constructed based on the empirical data, and some of the model specifications are rejected, leaving a set of model specifications. Finally, the common dynamics shared by all these remaining model specifications are regarded as the model result.

For most ABMs, the calibration usually starts with the direct calibration methods. Then, for the rest parameters that cannot be directly calibrated, the simulation-based calibration methods could be applied, as they imply the golden criterion of model validity: the consistency between the model output and the real data of the system.

However, due to the lack of data at the agent level and the explosion of parameter combinations, both the direct and simulation-based calibration methods are limited to calibrating a smaller group of environment parameters or the parameters that are homogeneous for all the agents. For example, Barde and van der Hoog [60] estimated 8 out of 33 parameters for a large-scale economic ABM, the Eurace@Unibi model [61], and the behavioral parameters are homogeneous among the agents. Kukacka and Barunik [55] estimated the financial ABM developed by Brock and Hommes [7], and the behavioral parameters are also homogeneous for the agents.

As a result, the existing calibration methods are limited in the following situation: when the agents are heterogeneous regarding their endowments or objectives, they should also be calibrated individually with heterogeneous strategies (i.e., behavioral parameters). Besides, as shown in Table I, improving model validity also concerns improving the validity of "mechanisms." For the environment, the mechanism (notably of agents' interaction) highly depends on the context, while for the agents, as reviewed in Section II-A, there are different implementations for each of the three modules—cognitive architecture, decision formation, and learning and adaption. Existing methods focus on calibrating the parameters of a given structure (i.e., function form) and provide limited insight on the choice of structure.

Apart from parameter calibration, another strand of studies discussed the implication of learning and adaptation for the validity of ABMs, and it was first under the topic of "rational expectation (RE)." The question was: will learning lead the interaction between boundedly rational agents to a rational equilibrium, and if so, how fast could it be?

Arifovic [28] first applied the genetic algorithm (GA) in modeling agent' learning in the cobweb model and found that GA works better than the other frequently studied algorithms in: 1) converging to the RE equilibrium for a wider range of parameter values and 2) capturing several features of the experimental behavior of human subjects. Arifovic [28] also compared social and individual learning based on GA, which was further analyzed by several other studies [29], [62], [63]. In a following study, Arifovic and Ledyard [64] proposed the "individual evolutionary learning (IEL)" model for mechanism design, which was also used for correlated equilibrium formation [65] and market design [66], [67]. Most recently, Anufriev *et al.* [13] used a GA-based individual learning model to show how agents learn to use smart heuristics in a complex environment. The output dynamics were compared with the stylized facts of learning-to-forecast experiments. The results showed that the model captures individual forecasting behavior in the experiments quite well and also reproduces the aggregate outcomes. This further enhanced the implication of learning for model validity.

III. General Framework

Following the review in Section II, this section introduces the proposed framework for policy simulation in two parts. First, for modeling agents' behaviors (i.e., policy functions), the form of sigmoid function is used and modified for its flexibility to approximate different function forms (Section III-A). Second, evolutionary training is used to individually calibrate the behavioral parameters of agents (Section III-B).

A. Behavior Modeling: Modified Sigmoid Function

As reviewed in Section II-A, the design of agents includes three modules and each module can be implemented in different ways as long as they can fit together. For simplification and generality, this study chooses the sigmoid function [$f(x) = 1/(1+e^{-x})$] as the starting point to construct the agents' "policy function(s)" for its appropriate properties: 1) continuity and smoothness; 2) unlimited range of independent variable; and 3) limited value range within $(0, 1)$.

Taking a technology diffusion model as an example, we assume that the agents need to make technology adoption decisions based on its current state and available technology options to maximize its total profit by the end of the simulation. Thus, the agent's "tendency" to adopt one technology can be modeled as a "probability" by (1), in which the impacts of two factors are integrated by the modified sigmoid function: 1) cost of the technology (C), which has a negative impact on the adoption tendency, and 2) potential benefit of adoption (B), which has a positive impact on the adoption tendency

$$\mathbf{Pr} = \left(\frac{1}{1+{\delta_1}^{C}}\right)^{\delta_2}\left(\frac{1}{1+{\delta_3}^{-B}}\right)^{\delta_4}. \quad (1)$$

Given the property of the sigmoid function, both parts on the right-hand side have ranges of $(0, 1)$, so their product $\mathbf{Pr}$ on the left-hand side is also in $(0, 1)$, corresponding to its meaning as a probability. δ_n ($n = 1, 2, 3, 4$) are four behavioral parameters. δ_1 and δ_3 capture the sensitivity of the agent's adoption tendency to price and social influence. The monotonicities of the two impacts are guaranteed when δ_1 and δ_3 are larger than 1. δ_2 and δ_4 capture the relative weights of the two impacts.

Taking a modified form of the sigmoid function, (1) implements the "decision formation" module of the agents facing the repetitive decision on technology adoption: the concept of "utility" is jumped over, and the impacts of identified factors are linked to the "adoption tendency" directly. By adding more similar segments on the right-hand side, we can integrate the impacts of more factors flexibly, as long as we know their monotonicities. Besides, the meaning of the behavioral parameters is interpretable.

On the other hand, for the "cognitive architecture" and "learning and adaptation" modules, we can also construct the corresponding functions by modifying the sigmoid function to model agents' tendency, as a probability to adjust their forecasts or behaviors, which is influenced by a set of factors with known monotonicities. More examples are provided in Section IV.

B. Behavior Calibration: Evolutionary Training

As shown in (1), the values of behavioral parameters δ_n ($n = 1, 2, 3, 4$) decide the monotonicities of influencing factors, as well as the sensitivity and relative weights of their impacts. To guarantee the reasonability of agents' behaviors, as well as the validity of the model, it is important to find proper values for agents' behavioral parameters. Thus, as the second part of this policy simulation framework, evolutionary training is used to calibrate the behavioral parameters of agents. Specifically, before the "simulation stage" for the *ex ante* analysis of a policy, a "training stage" is implemented, including the three steps as follows.

1) Randomly generate the initial behavioral parameters for the agents and start the training stage.
2) Let the agents individually update their behavioral parameters based on the GA.
3) Stop the training stage according to specified criteria, for example, the convergence of agents' behavioral parameters at the micro level or the result variables at the macro level.

Conceptually, this evolutionary training method is different from the methods reviewed by Platt [49] for parameter calibration because the model validity is not empirically evaluated but only improved or disciplined by optimization. In other words, the performance requirement for the model validity is not "producing data that are close enough to the observed data," but specified as "the agents are smart enough to make reasonable decisions." This is a compromise when simulation-based calibration is not possible due to lack of empirical data. On the other hand, the evolutionary training method also allows individually calibrating agents' behavioral parameters to consider their heterogeneity regarding endowments or objectives.[6]

In summary, by combining the optimization and simulation methods, this framework balances the flexibility and validity of ABMs. For the cases when simulation-based calibration is not possible due to lack of data, we can develop "test-bed" models within this framework to evaluate the *ex ante* impact of policies. An example with details of implementation is provided in Section IV.

IV. Framework Application Example

This section introduces an agent-based model for the emission trading scheme (AMETS) as an example [8], focusing on firms' behaviors modeling and the evolutionary training process. Section IV-A introduces the background and structure of AMETS. Then, the details of agents' behavior modeling based on the modified sigmoid function and the implementation of evolutionary training are provided in Section IV-B and IV-C. Finally, the results that show how the evolutionary training improves the model validity are provided in Section IV-D.

A. Model Background

The ETS is a key policy instrument for controlling the greenhouse gas (GHG) emission [68]. At the beginning, the government sets a target for total emission and allocates the allowances to covered firms from different sectors. Then, in the abatement phase, the firms trade with each other and also make production adjustments and low-carbon technology adoption decisions. By coordinating these three options, the aim of the firms is to reach the abatement target at the lowest cost. At the macro level, the policy is to form price signal and promote low-carbon technologies diffusion and ultimately to mitigate the GHG emission at the lowest total social cost.

One key feature of the ETS is the heterogeneity among firms, not only from different sectors but also from the same sector. Three aspects of heterogeneity are included: 1) allowance gap, i.e., the difference between total emission and initially allocated allowances; 2) carbon intensity and product profit, which decides firms' abatement cost by reducing the output; and 3) cost and energy-saving effect of firms' available low-carbon technologies. These heterogeneities decide the liquidity and efficiency of the mechanism,

[6]For clarification, there are two points regarding the terms used in this article. First, this article uses "evolutionary training" instead of "evolutionary learning" to emphasize the different motivations of using algorithms such as GA in ABMs. Second, the "calibration" in this article is different from its usual practice (fitting to observed data), but since it also concerns "finding proper values for parameters," the term "calibration" is still used.

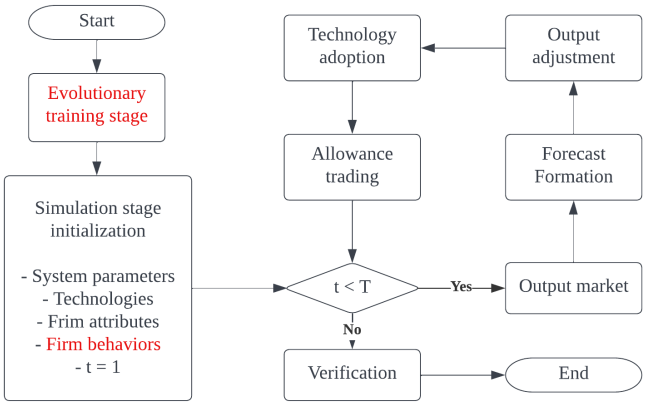

Fig. 2. Flowchart of AMETS.

and at the micro level, they also indicate firms' heterogeneous behaviors. For example, a firm will have a stronger tendency to reduce its output or buy allowances from others, if its available low-carbon technologies are more expensive.

In a previous paper [8], we established an AMETS. Compared with the other studies, AMETS contributes to the modeling of ETS by considering the following complexities: 1) different planning horizons of the three abatement options; 2) heterogeneity among sectors and firms; and 3) details of firms' production processes and low-carbon technologies. Based on the European data from the FORECAST model [69], AMETS is calibrated to cover five industrial sectors, 11 emission-intensive products, 25 production processes, and 52 low-carbon technologies. AMETS can be used as a policy simulation platform for the mechanism design and impact evaluation of the ETS. The flow diagram of AMETS is shown in Fig. 2.

At the beginning, AMETS starts with an evolutionary training stage to find firms' behavioral parameters. Then, combined with other data, the simulation stage is initialized for one abatement phase that lasts for T periods. In each period t, firms first compete in the output markets, where their profits and GHG emissions are decided. Then, firms will form their forecasts of the allowance price by observing the other firms and the allowance price in the past periods. Taking the forecast as a benchmark, firms will coordinate three abatement options by comparing their cost with the forecast and make decisions, including output adjustment, low-carbon technology adoption, and allowance trading. For the firms with cheaper technologies, they may adopt more than needed and sell allowances to the other firms, while for the firms with higher abatement cost, they will buy in the allowance market. By the end of the abatement phase, firms' total emissions are verified, and those who emit more than their holdings must pay a fine for each ton of excess emissions.

The fundamental complexity of the ETS is that firms coordinate three abatement options at the micro level, which then emerges into dynamic interactions among the allowance market, output markets, and low-carbon technology diffusion at the macro level. For example, when there comes a negative demand shock in the iron and steel market, the firms will reduce their output, as well as their demand of

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE III

IMPACTS OF RELEVANT FACTORS ON FIRM i'S DECISION TENDENCIES

Factor	Explanation	T1	T2	T3
$FP_{i,t}^{A}$	forecast of the allowance price	+	+	+
$ENA_{i,t}$	expected net allowance surplus	−	−	−
$PUE_{i,t}^{O}$	profit of unit emission in the output market	−	○	○
$AC_{i,j}$	average abatement cost of technology j	○	−	○
$AAP_{i,j,t}$	absolute abatement potential of technology j	○	+	○
$P_{t,k}^{A}$	allowance price at the current tick k	○	○	−

emission allowances. Then, this may bring down the allowance price, which may further increase the production and slow down the technology diffusion in the other sectors.

B. Behavior Modeling

In AMETS, firms coordinate the three abatement options by comparing their costs with a benchmark, i.e., firms' forecast of the allowance price. The abatement costs of the three options are given as follows.

1) *Output Adjustment:* Profit of unit emission in the output market.
2) *Technology Adoption:* Average abatement cost of the low-carbon technology.
3) *Allowance Trading:* Current allowance price.

Besides, tendencies for choosing the three options are also influenced by firms' expected net allowance surplus, which is the difference between holdings of allowances and expected total emissions. Finally, when considering to adopt a low-carbon technology, the tendency is also influenced by the technologies' absolute abatement potential in the remaining periods.

Table III summarizes the impacts of relevant factors on firm i's three tendencies. "+" and "−" indicate positive and negative impacts, respectively. Second-order impacts are neglected and marked with ○ because firms go through the three decisions in every period. Once a decision is made in one period, it will immediately influence the other decisions from the next period on.

In order to synthesize the impacts of multiple factors on firms' tendencies for three decisions, three sets of behavioral functions are designed based on the sigmoid function.

First, for the output adjustment decision, each firm i will first calculate a threshold price of allowance ($\mathrm{TP}_{i,t}^{O}$), which increases with the increase of allowance price forecast ($\mathrm{FP}_{i,t}^{A}$) and negatively influenced by the expected net allowance surplus ($\mathrm{ENA}_{i,t}$), as shown in (2), where e_i is firm i's daily CO_2 emissions at the beginning of simulation. Then, firm i will compare its profit of unit emission in the output market ($\mathrm{PUE}_{i,t}^{O}$) with this threshold and decides to increase or decrease its output, if $\mathrm{PUE}_{i,t}^{O}$ is higher or lower than $\mathrm{TP}_{i,t}^{O}$. The two probabilities are $\Phi_{i,t}^{1}$ or $\Phi_{i,t}^{2}$, calculated by (3) and (4), respectively

$$\mathrm{TP}_{i,t}^{O} = \mathrm{FP}_{i,t}^{A} \cdot \left[1 + \alpha_{i,1}\left(\frac{1}{1+\alpha_{i,2}{}^{\mathrm{ENA}_{i,t}/e_i}} - 0.5\right)\right] \quad (2)$$

TABLE IV

VALUE RANGES OF FIRMS' BEHAVIORAL PARAMETERS

Parameter	[min, max]	Parameter	[min, max]	Parameter	[min, max]
$\alpha_{i,1}$	[0,2]	$\beta_{i,1}$	[1,5]	$\gamma_{i,1}$	[0,2]
$\alpha_{i,2}$	[1,5]	$\beta_{i,2}$	[0.25,4]	$\gamma_{i,2}$	[1,5]
$\alpha_{i,3}$	[1,5]	$\beta_{i,3}$	[1,5]	$\gamma_{i,3}$	[1,5]
$\alpha_{i,4}$	[1,5]	$\beta_{i,4}$	[0.25,4]	$\gamma_{i,4}$	[1,5]

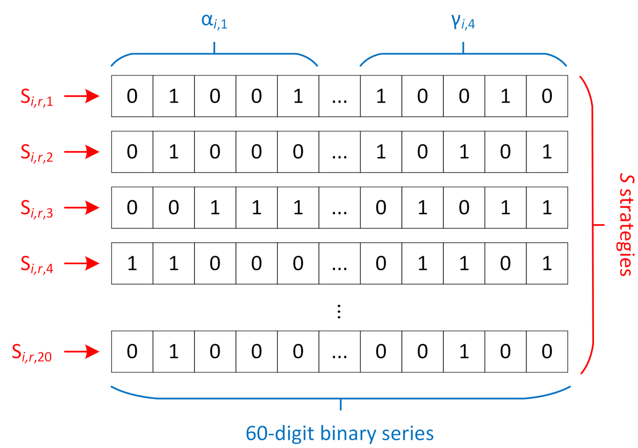

Fig. 3. Firm i's population of strategies in the rth generation.

$$\Phi_{i,t}^{1} = 2 \times \left(0.5 - \frac{1}{1+\alpha_{i,3}{}^{\mathrm{PUE}_{i,t}^{O}/\mathrm{TP}_{i,t}^{O}}}\right) \quad (3)$$

$$\Phi_{i,t}^{2} = 2 \times \left(0.5 - \frac{1}{1+\alpha_{i,4}{}^{\mathrm{TP}_{i,t}^{O}/\mathrm{PUE}_{i,t}^{O}}}\right). \quad (4)$$

Second, for the low-carbon technology adoption decision, each firm i's tendency to adopt technology j is modeled as a probability ($\Psi_{i,t}$) and decided by the four factors summarized in Table III, as calculated by the following equation:

$$\Psi_{i,t} = \left(\frac{1}{1+\beta_{i,1}{}^{\mathrm{AC}_{i,j}/\mathrm{FP}_{i,t}^{A}}}\right)^{\beta_{i,2}} \left(\frac{1}{1+\beta_{i,3}{}^{\mathrm{ENA}_{i,t}/\mathrm{AAP}_{i,j,t}}}\right)^{\beta_{i,4}}. \quad (5)$$

Third, for the allowance trading decision, each firm i will follow similar procedures for output adjustment decision, calculating a threshold price of allowance ($\mathrm{TP}_{i,t}^{A}$), as shown in (6). Then, firm i will compare the current allowance price ($P_{t,k}^{A}$)[7] with this threshold and decides to sell or buy allowance with a probability of $\Omega_{i,t}^{1}$ or $\Omega_{i,t}^{2}$, calculated by (7) or (8), respectively

$$\mathrm{TP}_{i,t}^{A} = \mathrm{FP}_{i,t}^{A} \cdot \left[1 + \gamma_{i,1}\left(\frac{1}{1+\gamma_{i,2}{}^{\mathrm{ENA}_{i,t}/e_i}} - 0.5\right)\right] \quad (6)$$

$$\Omega_{i,t,k}^{1} = 2 \times \left(0.5 - \frac{1}{1+\gamma_{i,3}{}^{P_{t,k}^{A}/\mathrm{TP}_{i,t}^{A}}}\right) \quad (7)$$

$$\Omega_{i,t,k}^{2} = 2 \times \left(0.5 - \frac{1}{1+\gamma_{i,4}{}^{\mathrm{TP}_{i,t}^{A}/P_{t,k}^{A}}}\right). \quad (8)$$

[7] In AMETS, the allowance trading process is modeled based on the continuous double auction mechanism. Each period t is further divided into N_K ticks, number by k. On each tick k, the allowance price is $P_{t,k}^{A}$, and a firm is selected to make a trading decision. After the matching process, firms' allowance and cash accounts are updated.

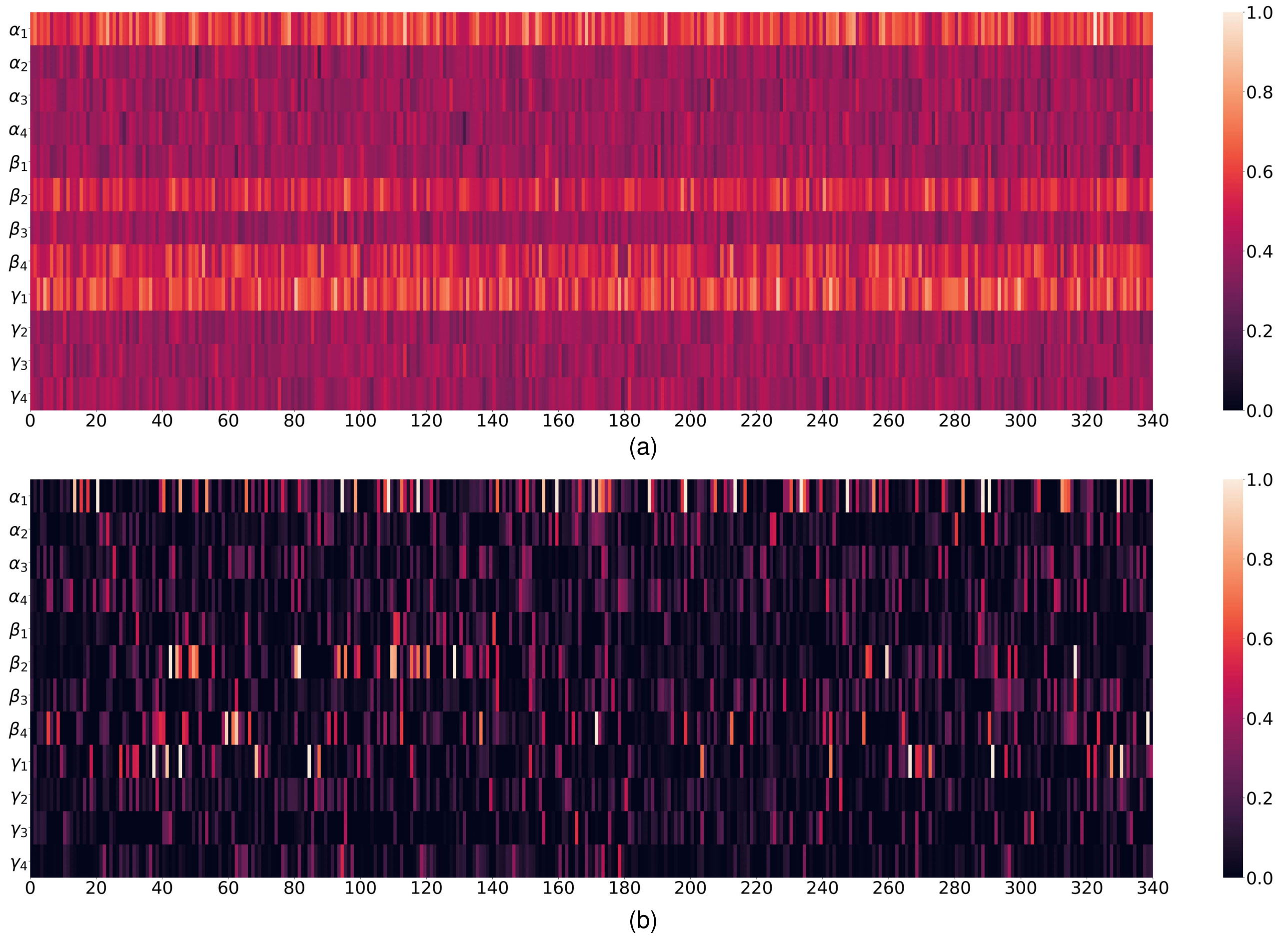

Fig. 4. CoV (color) of 12 × 340 = 4080 behavioral parameters of all firms: convergence in the training stage. (a) Start of the training stage. (b) End of the training stage.

In (2)–(8), there are three sets of behavioral parameters: 1) $\alpha_{i,l}$ ($l = 1, 2, \ldots, 4$) for output adjustment decision; 2) $\beta_{i,m}$ ($m = 1, 2, \ldots, 4$) for technology adoption decision; and 3) $\gamma_{i,n}$ ($n = 1, 2, \ldots, 4$) for allowance trading decision. To guarantee the factors influencing agents' three decisions as summarized in Table III, the 12 behavioral parameters are set in specific ranges, as shown in Table IV.

C. Evolutionary Training

To calibrate the behavioral parameters of the firms, an evolutionary training stage is introduced based on the GA. Following the common GA setup, values of the 12 behavioral parameters are coded as a 60-digit binary series, which is the "chromosome." At the beginning of the training stage, each firm i is randomly initialized with a population of 20 strategies.

As shown in Fig. 3, $S_{i,r,j}$ denotes the jth strategy of firm i in the rth generation. For each chromosome, every five-digit segment represents the value of one parameter. Taking $\alpha_{i,1}$ in $S_{i,r,1}$ (represented by "01001") as an example, its mapping to the parameter value, a decimal number in $[0, 2]$, includes two steps. First, map the five-digit binary number to a decimal number and divide it by 31, which is the biggest decimal number that a five-digit binary number can represent, and then, we get a number x in range $[0, 1]$. Thus, we have: $01001 \rightarrow 1 \cdot 2^3 + 1 \cdot 2^0 = 9 \rightarrow x = 9/31$. Second, for different behavioral parameters, we linearly map x to the decimal number in its range. Thus, we have $\alpha_{i,1} = 0 + 2x = 18/31$.

The evolutionary training stage runs as follows.

First, the training stage starts by randomly initializing firms' first generation of strategies. In each rth generation of training, the simulation of one abatement phase is run 20 times. In each time, each firm i will try one strategy j in its current population and record the final total profit ($\mathrm{FTP}_{i,r,j}$) by the end of the abatement phase.

Second, by the end of rth generation of training, firms will individually update their population of strategies based on GA and generate their new populations of strategies for the $(r+1)$th generation of training. Taking firm i as an example, the implementation of GA includes selection, crossover, and mutation as follows.

Third, as the training stage proceeds, firms' behavioral parameters at the micro level and the market results at the macro level converge, which indicates that the training stage comes to the end. Based on my test, the total number of generations is set to be 50, and the results of convergence are presented in Section IV-D. In the end, each firm i will choose the strategy with the highest fitness in the last generation as its strategy for the simulation stage afterward. By introducing

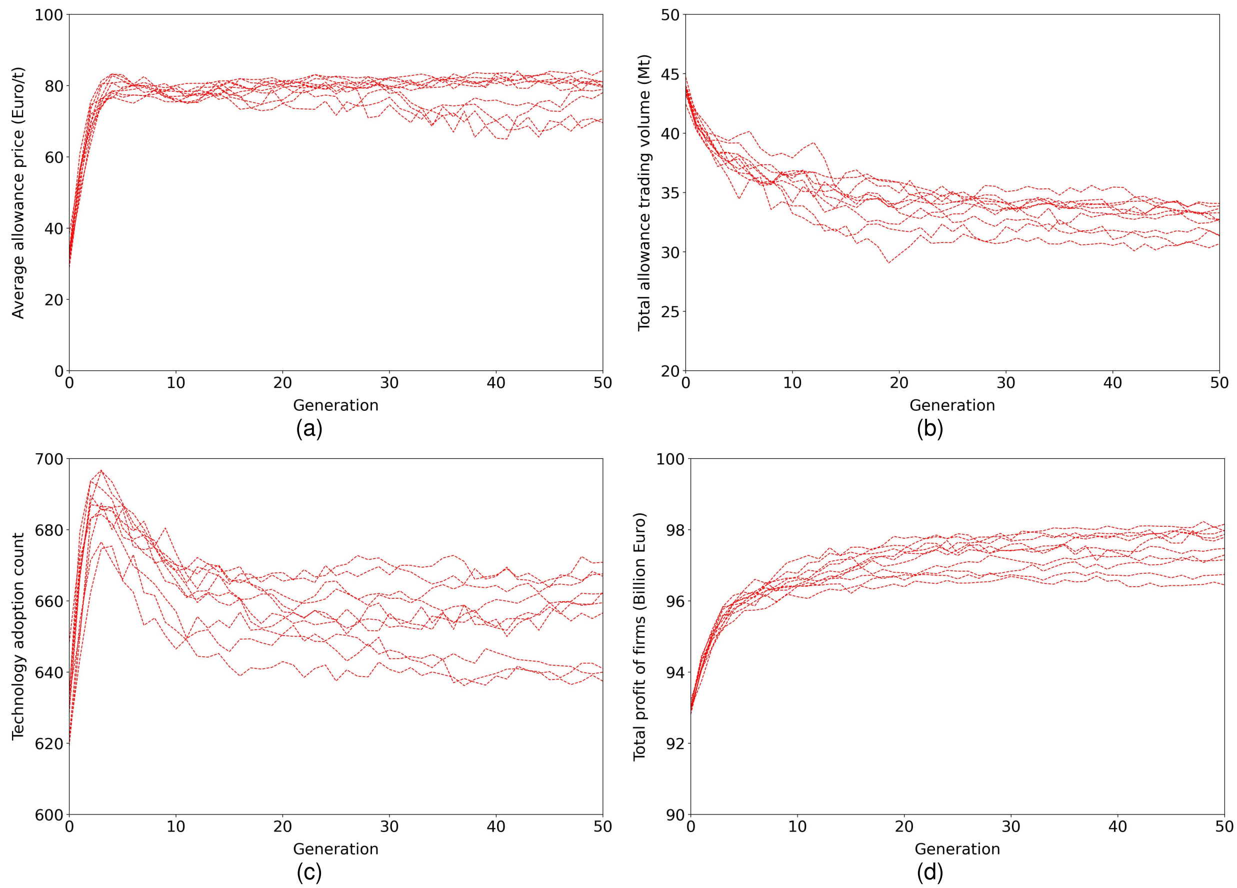

Fig. 5. Market results convergence in the training stage. (a) Average allowance price. (b) Total allowance trading volume. (c) Technology adoption count. (d) Total profit of firms.

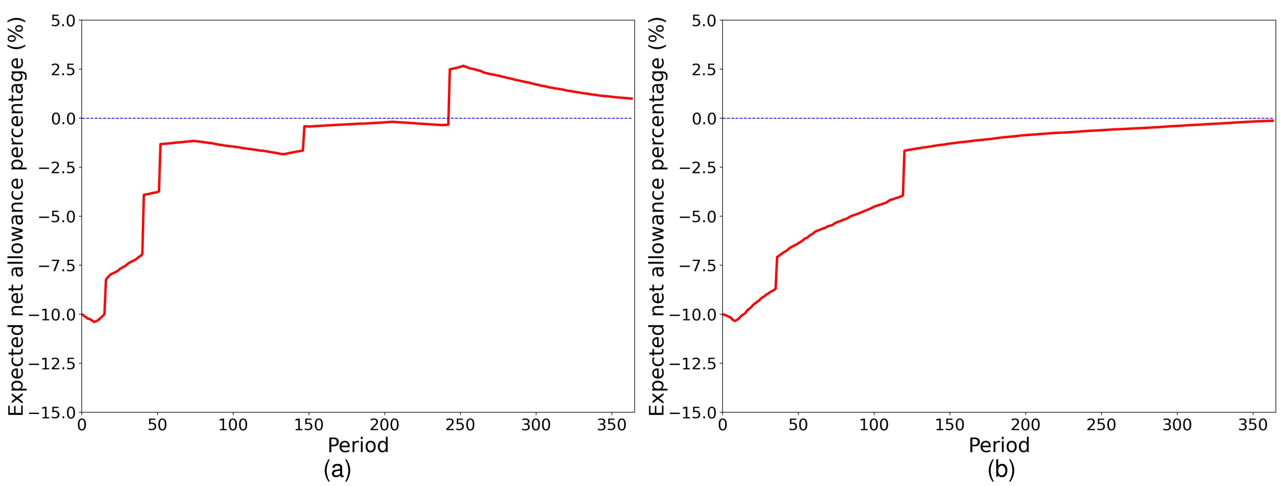

Fig. 6. Comparison between the expected net allowance of $\mathrm{BFSteel}_1$ under (a) WoTS and (b) WTS scenarios.

such an individual training stage based on GA, the behavioral parameters for 340 heterogeneous firms are found,[8] i.e., $12 \times 340 = 4080$ parameters in total.

[8]Based on the European data, AMETS is calibrated to cover 11 emission-intensive products, which are further divided into 17 because of different production processes. For each product, we set that there are 20 firms, so 340 firms in total.

D. Results

In this section, we focus on the results showing the impact of evolutionary training on the validity of AMETS, which are reflected in two parts: 1) convergence of firms' behavioral parameters and the market results in the training stage (Section IV-D1) and 2) comparison between the simulation results with and without the training stage (Section IV-D2).

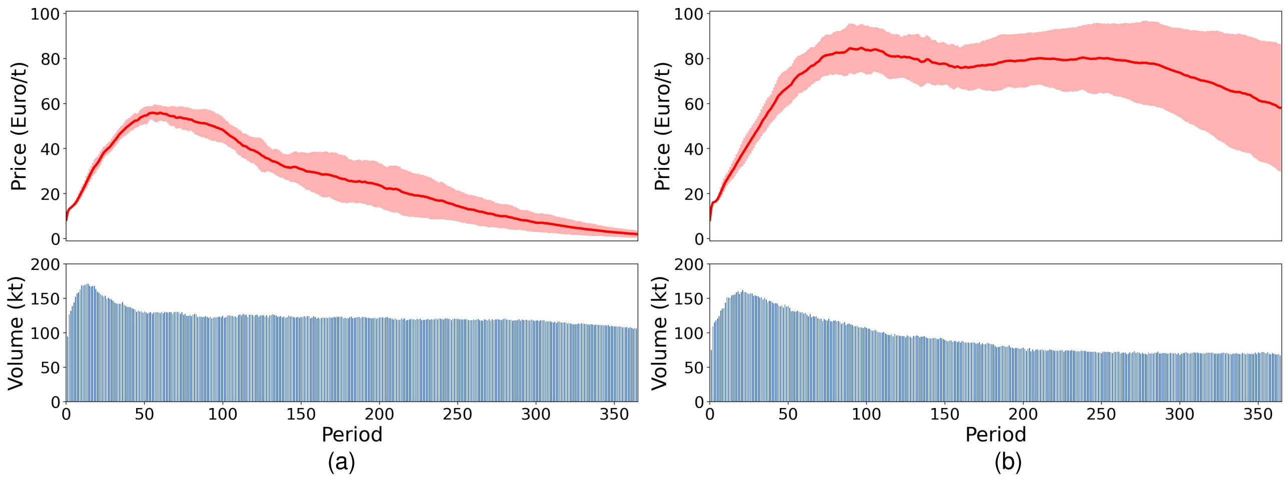

Fig. 7. Comparison between the price patterns in the allowance market under (a) WoTS and (b) WTS scenarios.

1) Convergence of Firms' Behavioral Parameters and Market Results: As adopted by previous studies since Arifovic [28], a straightforward criterion implying validity of "learning (i.e., training)" is convergence.

In the training stage of AMETS, firms individually update their 12 behavioral parameters based on GA. For each firm, in one generation of its strategies, there are 20 values for each behavioral parameter, and the variation of one parameter can be measured by the coefficient of variation (CoV), i.e., the ratio between its standard deviation and mean value. Thus, the convergence of firms' behavioral parameters can be represented by the decline of CoV.

Fig. 4 shows the convergence of all 4080 behavioral parameters by comparing their CoV at the start and end of the learning stage. Each rectangle represents one of the 4080 parameters, and its color represents the CoV value: darker color indicates a lower CoV value. After the training stage, most behavioral parameters have significantly converged (CoV_{mean} declined from 0.44 to 0.10).

At the macro level, training also leads to the convergence of market results. Four aggregate results are selected as examples, including average allowance price, total allowance trading volume, technology adoption count, and total profit of firms. As shown in Fig. 5, the training stage is run ten times with randomly initialized firms' seed strategies each time. After 50 generations, all the paths show significant convergence. Besides, the ten evolution paths follow a similar pattern, which indicates the robustness of the evolutionary training method. As shown in Fig. 5(d), firms improved their total profit in the training stage.

2) Comparison Between the Simulation Results With and Without the Training Stage: Apart from the convergence, the implication of the training stage for model validity is also reflected by comparing the simulation results under two scenarios, in which firms' behavioral parameters are: 1) randomly assigned without the training stage and 2) initialized with the training stage. The two scenarios are hereafter referred to as without training stage (WoTS) and with training stage (WTS).

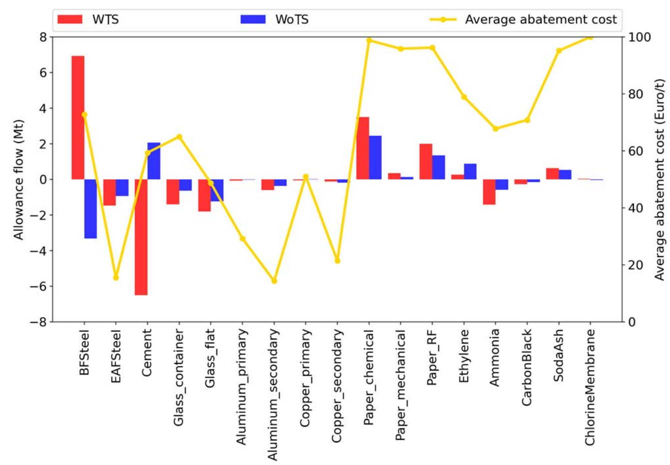

Fig. 8. Comparison between the allowance flow among sectors under WoTS and WTS scenarios.

As introduced in Section IV-A, firms in the ETS coordinate three options for GHG emission abatement, including output adjustment, low-carbon technology adoption, and allowance trading. By the end of the abatement phase, a firm will pay a fine for each ton of excess emissions if it emitted more than the allowances in its account. In AMETS, we assume that the allowance cannot be used in the next phase. Then, a smart firm will dynamically coordinate the three options, balance its expected net allowance and the allowance in the account, and try to exactly comply with the abatement target by the end of the phase. Thus, we select the expected net allowance surplus ($\text{ENA}_{i,t}$) of firms to show the rationality of their strategies. Taking one specific firm (BFSteel_1[9]) as an example, the evolution of its $\text{ENA}_{i,t}$ in the abatement phase is presented in Fig. 6.

As shown in Fig. 6, the firm started the abatement phase with a deficit of allowance, which represents the abatement target of 10%. Then, each sharp increase of the curve indicates the firm adopting one low-carbon technology. Under WoTS,

[9]Here, we selected the first firm in the blast furnace steel sector.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

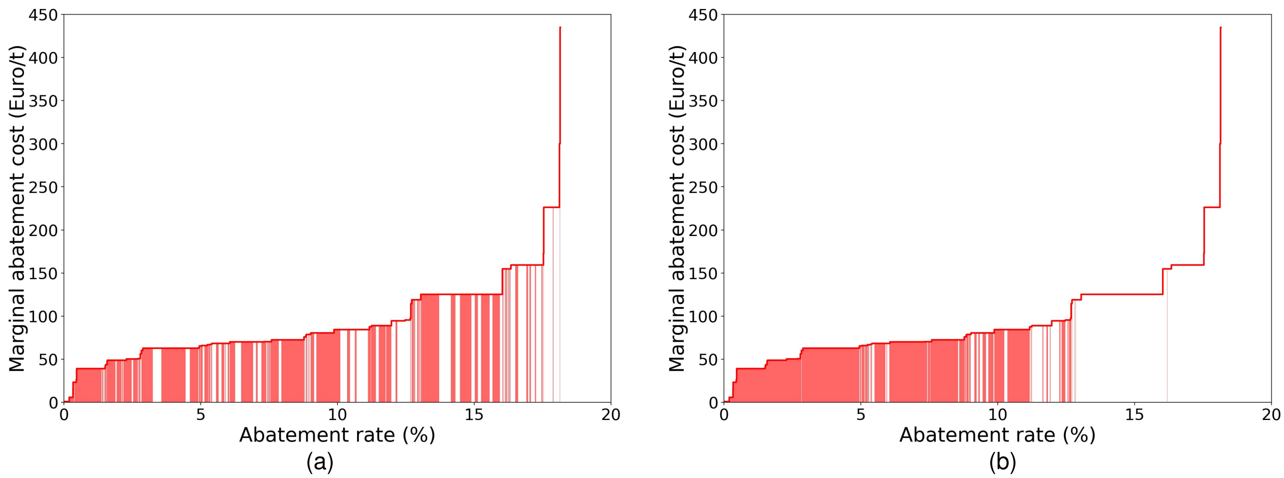

Fig. 9. Comparison between the technology diffusion under (a) WoTS and (b) WTS scenarios.

the firm adopted five low-carbon technologies, and the last one led to a significant surplus of the allowance, while under WTS, the firm adopted two, meaning lower investment cost than WoTS. By the end of the abatement phase, the firm left more allowance in the account under WoTS than WTS. In the end, the firm earned higher total profit under WTS than WoTS.

The implication of the training stage is also reflected by market results at the macro level.

First, in the allowance market, we observe different price patterns under the two scenarios. As shown in Fig. 7, when the abatement target is set at 10%, the allowance price is higher under WTS than WoTS. The red line represents the mean of ten runs, and the shadow represents the range of one standard deviation. In the training stage, firms learned the pressure of an abatement target at 10% and become more cautious about selling allowance.

Second, regarding the trading volume, firms with higher abatement cost will buy allowances from those whose abatement costs are lower. Thus, we observe the allowance flow among sectors under the two scenarios, as shown in Fig. 8. Positive allowance flow indicates buyers in the market and negative allowance flow indicates sellers. Under WTS, the largest buyer is BFSteel (blast furnace steel sector) and the largest seller is Cement. As shown by the golden line, BFSteel has a higher abatement cost than Cement. However, under WoTS, the allowance flow between the two sectors is opposite. Besides, as shown in Fig. 8, for the sectors with the same trading position under the two scenarios, the absolute trading volume is higher under WTS than WoTS, which means that firms are clearer about their positions under WTS and trade more sufficiently.

Finally, higher overall efficiency of ETS under WTS is also observed in low-carbon technology diffusion. As shown in Fig. 9, all firms' available low-carbon technologies are ranked according to their average abatement cost, from low to high, as the marginal abatement cost curve (MACC) of the system. Each step of the curve represents one technology option. The shadowed parts indicate the technologies that are adopted by the end of the abatement phase. Under WoTS, more technologies with higher abatement cost are adopted leaving the cheaper ones not adopted, indicating lower efficiency than WTS.

V. Conclusion

Compared with the mainstream neoclassical economic models, ABMs have been criticized for their arbitrariness, specifically, the lack of general framework for modeling agents' behaviors, and the difficulties in ensuring the model validity. This article tries to respond to both of the questions by proposing a framework with two pillars: 1) using the sigmoid function as a building block to flexibly construct agents' decision-making rules and 2) using the evolutionary training method to calibrate agents' behavioral parameters and to improve the model validity. By combining the simulation and optimization methods, this framework balances the flexibility and validity of ABMs and contributes to the literature on ABM design, calibration, and application of the learning techniques.

This framework can be applied to develop "test-bed" models for policy simulation when two conditions are satisfied: 1) the agents try to maximize some intertemporal preference and 2) the impacts of different factors on agents' behavioral tendency are monotonic. Without relying on agent-level empirical data, the model validity is supported by direct calibration and evolutionary training.

Furthermore, this framework can still be coupled with a simulation-based calibration method when empirical data of the target system are available. The "evolutionary training" stage is essentially "solving" the model. When coupled with the simulation-based calibration method, the overall framework becomes comparable with the structural estimation framework in the mainstream economics.

This framework has two limitations: 1) the two conditions for using the framework limit its application and 2) the "training stage" can demand long running time if the model is complicated. Besides, when applying the framework, modelers should also be aware of the "overtraining" problem, i.e., the agents could be trained to be "too smart." In some cases, it could be helpful to stop the training stage earlier or leave part of the agents not trained.

To support the framework validity, further studies and results comparison are helpful: 1) applying the framework to other modeling cases; 2) using other function forms with similar properties of the sigmoid function; and 3) using other evolutionary algorithms other than GA [e.g., particle swarm optimization (PSO)] for training the behavioral parameters.

References

[1] T. Leigh, "Agent-based computational economics: A constructive approach to economic theory," in *Handbook of Computational Economics*, vol. 2. Amsterdam, The Netherlands: Elsevier, 2006, pp. 831–880.

[2] G. Iori, "A microsimulation of traders activity in the stock market: The role of heterogeneity, agents' interactions and trade frictions," *J. Econ. Behav. Org.*, vol. 49, no. 2, pp. 269–285, 2002.

[3] A. Trivedi and S. Rao, "Agent-based modeling of emergency evacuations considering human panic behavior," *IEEE Trans. Computat. Social Syst.*, vol. 5, no. 1, pp. 277–288, Mar. 2018.

[4] M. Raberto and S. Cincotti, "Modeling and simulation of a double auction artificial financial market," *Phys. A, Stat. Mech. Appl.*, vol. 355, no. 1, pp. 34–45, Sep. 2005.

[5] S.-H. Chen and U. Gostoli, "Coordination in the El Farol bar problem: The role of social preferences and social networks," *J. Econ. Interact. Coordination*, vol. 12, no. 1, pp. 59–93, Apr. 2017.

[6] H. Zhu, "Computational social simulation with E-CARGO: Comparison between collectivism and individualism," *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 6, pp. 1345–1357, Mar. 2020.

[7] W. A. Brock and C. H. Hommes, "Heterogeneous beliefs and routes to chaos in a simple asset pricing model," *J. Econ. Dyn. Control*, vol. 22, nos. 8–9, pp. 1235–1274, Jul. 1998.

[8] S.-M. 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.*, vol. 286, no. 3, pp. 1113–1128, 2020.

[9] W. B. Arthur, J. H. Holland, B. LeBaron, R. Palmer, and P. Tayler, "Asset pricing under endogenous expectations in an artificial stock market," in *The Economy as an Evolving Complex System II*. Tallahassee, FL, USA: CRC Press, 2018, pp. 15–44. [Online]. Available: https://www.taylorfrancis.com/chapters/edit/10.1201/9780429496639-2/asset-pricing-endogenous-expectations-artificial-stock-market-brian-arthur-john-holland-blake-lebaron-richard-palmer-paul-tayler

[10] D. K. Gode and S. Sunder, "Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality," *J. Political Economy*, vol. 101, no. 1, pp. 119–137, Feb. 1993.

[11] J. Conlisk, "Why bounded rationality?" *J. Econ. Literature*, vol. 34, no. 2, pp. 669–700, 1996.

[12] J. Q. Dong, "Technological choices under uncertainty: Does organizational aspiration matter?" *Strategic Manage. J.*, vol. 42, no. 5, pp. 898–916, May 2021.

[13] M. Anufriev, C. Hommes, and T. Makarewicz, "Simple forecasting heuristics that make us smart: Evidence from different market experiments," *J. Eur. Econ. Assoc.*, vol. 17, no. 5, pp. 1538–1584, 2019.

[14] T. Bao, C. Hommes, and J. Pei, "Expectation formation in finance and macroeconomics: A review of new experimental evidence," *J. Behav. Exp. Finance*, vol. 32, Dec. 2021, Art. no. 100591.

[15] M. Le Pira, E. Marcucci, V. Gatta, G. Inturri, M. Ignaccolo, and A. Pluchino, "Integrating discrete choice models and agent-based models for ex-ante evaluation of stakeholder policy acceptability in urban freight transport," *Res. Transp. Econ.*, vol. 64, pp. 13–25, Sep. 2017.

[16] B. LeBaron, "Agent-based computational finance," in *Handbook of Computational Economics*, vol. 2. Amsterdam, The Netherlands: Elsevier, 2006, pp. 1187–1233.

[17] E. J. Rykiel, "Testing ecological models: The meaning of validation," *Ecol. Model.*, vol. 90, no. 3, pp. 229–244, Nov. 1996.

[18] G. Gigerenzer, "Fast and frugal heuristics: The tools of bounded rationality," in *Blackwell Handbook of Judgment and Decision Making*, vol. 62. Hoboken, NJ, USA: Wiley, 2004, p. 88.

[19] J. Cyert and R. March, *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1963.

[20] H. R. Greve, *Organizational Learning from Performance Feedback: A Behavioral Perspective on Innovation and Change*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[21] V. Gaba and J. Joseph, "Corporate structure and performance feedback: Aspirations and adaptation in M-form firms," *Org. Sci.*, vol. 24, no. 4, pp. 1102–1119, Aug. 2013.

[22] D. McFadden, "The measurement of urban travel demand," *J. Public Econ.*, vol. 3, no. 4, pp. 303–328, 1974.

[23] R. G. Palmer, W. B. Arthur, J. H. Holland, B. LeBaron, and P. Tayler, "Artificial economic life: A simple model of a stock market," *Phys. D, Nonlinear Phenomena*, vol. 75, nos. 1–3, pp. 264–274, 1994.

[24] P. Albin and D. K. Foley, "Decentralized, dispersed exchange without an auctioneer: A simulation study," *J. Econ. Behav. Org.*, vol. 18, no. 1, pp. 27–51, 1992.

[25] J. Arifovic, B. C. Eaton, and G. Walker, "The coevolution of beliefs and networks," *J. Econ. Behav. Org.*, vol. 120, pp. 46–63, Dec. 2015.

[26] T. Lux, "Herd behaviour, bubbles and crashes," *Econ. J.*, vol. 105, no. 431, pp. 881–896, 1995.

[27] A. Kirman, "Ants, rationality, and recruitment," *Quart. J. Econ.*, vol. 108, no. 1, pp. 137–156, Feb. 1993.

[28] J. Arifovic, "Genetic algorithm learning and the cobweb model," *J. Econ. Dyn. Control*, vol. 18, no. 1, pp. 3–28, Jan. 1994.

[29] N. J. Vriend, "An illustration of the essential difference between individual and social learning, and its consequences for computational analyses," *J. Econ. Dyn. Control*, vol. 24, no. 1, pp. 1–19, Jan. 2000.

[30] A. Beltratti *et al.*, *Neural Networks for Economic and Financial Modelling*. London, U.K.: International Thomson Computer Press, 1996.

[31] P. Ye, T. Wang, and F.-Y. Wang, "A survey of cognitive architectures in the past 20 years," *IEEE Trans. Cybern.*, vol. 48, no. 12, pp. 3280–3290, Aug. 2018.

[32] P. J. Ye, S. Wang, and F.-Y. Wang, "A general cognitive architecture for agent-based modeling in artificial societies," *IEEE Trans. Computat. Social Syst.*, vol. 5, no. 1, pp. 176–185, Mar. 2018.

[33] T. Brenner, "Agent learning representation: Advice on modelling economic learning," in *Handbook of Computational Economics*, vol. 2. Amsterdam, The Netherlands: Elsevier, 2006, pp. 895–947.

[34] W. B. Arthur, "Designing economic agents that act like human agents: A behavioral approach to bounded rationality," *Amer. Econ. Rev.*, vol. 81, no. 2, pp. 353–359, May 1991.

[35] O. Kirchkamp, "Spatial evolution of automata in the prisoners' dilemma," *J. Econ. Behav. Org.*, vol. 43, no. 2, pp. 239–262, 2000.

[36] B. LeBaron, "Empirical regularities from interacting long- and short-memory investors in an agent-based stock market," *IEEE Trans. Evol. Comput.*, vol. 5, no. 5, pp. 442–455, Oct. 2001.

[37] S.-H. Chen and C.-H. Yeh, "Evolving traders and the business school with genetic programming: A new architecture of the agent-based artificial stock market," *J. Econ. Dyn. Control*, vol. 25, nos. 3–4, pp. 363–393, Mar. 2001.

[38] T. Lux and M. Marchesi, "Volatility clustering in financial markets: A microsimulation of interacting agents," *Int. J. Theor. Appl. Finance*, vol. 3, no. 4, pp. 675–702, 2000.

[39] R. Selten, "Features of experimentally observed bounded rationality," *Eur. Econ. Rev.*, vol. 42, nos. 3–5, pp. 413–436, May 1998.

[40] T. Brenner and U. Witt, "Melioration learning in games with constant and frequency-dependent pay-offs," *J. Econ. Behav. Org.*, vol. 50, no. 4, pp. 429–448, Apr. 2003.

[41] H. D. Dixon, "Keeping up with the joneses: Competition and the evolution of collusion," *J. Econ. Behav. Org.*, vol. 43, no. 2, pp. 223–238, Oct. 2000.

[42] T. Börgers and R. Sarin, "Learning through reinforcement and replicator dynamics," *J. Econ. Theory*, vol. 77, no. 1, pp. 1–14, Nov. 1997.

[43] V. P. Crawford, "Adaptive dynamics in coordination games," *Econ., J. Econ. Soc.*, vol. 63, no. 1, pp. 103–143, Jan. 1995.

[44] R. Chakrabarti and R. Roll, "Learning from others, reacting, and market quality," *J. Financial Markets*, vol. 2, no. 2, pp. 153–178, May 1999.

[45] A. Marcet and T. J. Sargent, "Convergence of least squares learning mechanisms in self-referential linear stochastic models," *J. Econ. Theory*, vol. 48, no. 2, pp. 337–368, Aug. 1989.

[46] S. Manson, "Validation and verification of multi-agent models for ecosystem management," in *Complexity and Ecosystem Management: The Theory and Practice of Multi-Agent Approaches*. Cheltenham, U.K.: Edward Elgar, 2003, pp. 63–74.

[47] C. Bianchi, P. Cirillo, M. Gallegati, and P. A. Vagliasindi, "Validating and calibrating agent-based models: A case study," *Comput. Econ.*, vol. 30, no. 3, pp. 245–264, Sep. 2007.

[48] S. Alfarano, T. Lux, and F. Wagner, "Estimation of agent-based models: The case of an asymmetric herding model," *Comput. Econ.*, vol. 26, no. 1, pp. 19–49, Aug. 2005.

[49] D. Platt, "A comparison of economic agent-based model calibration methods," *J. Econ. Dyn. Control*, vol. 113, Apr. 2020, Art. no. 103859.

[50] M. Gilli and P. Winker, "A global optimization heuristic for estimating agent based models," *Comput. Statist. Data Anal.*, vol. 42, no. 3, pp. 299–312, Mar. 2003.

[51] C. Gourieroux, A. Monfort, and E. Renault, "Indirect inference," *J. Appl. Econ.*, vol. 8, pp. S85–S118, Dec. 1993.

[52] M. C. Recchioni, G. Tedeschi, and M. Gallegati, "A calibration procedure for analyzing stock price dynamics in an agent-based framework," *J. Econ. Dyn. Control*, vol. 60, pp. 1–25, Nov. 2015.

[53] S. Barde, "A practical, accurate, information criterion for Nth order Markov processes," *Comput. Econ.*, vol. 50, no. 2, pp. 281–324, Aug. 2017.

[54] F. Lamperti, "An information theoretic criterion for empirical validation of simulation models," *Econ. Statist.*, vol. 5, pp. 83–106, Jan. 2018.

[55] J. Kukacka and J. Barunik, "Estimation of financial agent-based models with simulated maximum likelihood," *J. Econ. Dyn. Control*, vol. 85, pp. 21–45, Dec. 2017.

[56] T. Lux, "Estimation of agent-based models using sequential Monte Carlo methods," *J. Econ. Dyn. Control*, vol. 91, pp. 391–408, Jun. 2018.

[57] J. Grazzini, M. G. Richiardi, and M. Tsionas, "Bayesian estimation of agent-based models," *J. Econ. Dyn. Control*, vol. 77, pp. 26–47, Apr. 2017.

[58] C. Werker and T. Brenner, "Empirical calibration of simulation models," Eindhoven Center Innov. Stud., Eindhoven, The Netherlands, Working Paper 04.13, 2004. [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.632.3809&rep=rep1&type=pdf

[59] G. Fagiolo, M. Guerini, F. Lamperti, A. Moneta, and A. Roventini, "Validation of agent-based models in economics and finance," in *Computer Simulation Validation*. Berlin, Germany: Springer, 2019, pp. 763–787.

[60] S. Barde and S. Van Der Hoog, "An empirical validation protocol for large-scale agent-based models," School Econ., Univ. Kent, Canterbury, U.K., Discuss. Papers 1712, 2017. [Online]. Available: https://www.econstor.eu/bitstream/10419/189902/1/1712.pdf

[61] H. Dawid, P. Harting, S. Van Der Hoog, and M. Neugart, "Macroeconomics with heterogeneous agent models: Fostering transparency, reproducibility and replication," *J. Evol. Econ.*, vol. 29, no. 1, pp. 467–538, Mar. 2019.

[62] J. Arifovic and M. K. Maschek, "Revisiting individual evolutionary learning in the cobweb model—An illustration of the virtual spite-effect," *Comput. Econ.*, vol. 28, no. 4, pp. 333–354, 2006.

[63] T. Vallée and M. Yildizoğlu, "Convergence in the finite cournot oligopoly with social and individual learning," *J. Econ. Behav. Org.*, vol. 72, no. 2, pp. 670–690, Nov. 2009.

[64] J. Arifovic and J. Ledyard, "A behavioral model for mechanism design: Individual evolutionary learning," *J. Econ. Behav. Org.*, vol. 78, no. 3, pp. 374–395, May 2011.

[65] J. Arifovic, J. F. Boitnott, and J. Duffy, "Learning correlated equilibria: An evolutionary approach," *J. Econ. Behav. Org.*, vol. 157, pp. 171–190, Jan. 2019.

[66] M. Anufriev, J. Arifovic, J. Ledyard, and V. Panchenko, "Efficiency of continuous double auctions under individual evolutionary learning with full or limited information," *J. Evol. Econ.*, vol. 23, no. 3, pp. 539–573, Jul. 2013.

[67] M. Van De Leur and M. Anufriev, "Timing under individual evolutionary learning in a continuous double auction," *J. Evol. Econ.*, vol. 28, no. 3, pp. 609–631, Aug. 2018.

[68] R. G. Newell, W. A. Pizer, and D. Raimi, "Carbon markets: Past, present, and future," *Annu. Rev. Resource Econ.*, vol. 6, no. 1, pp. 191–215, Nov. 2014.

[69] T. Fleiter *et al.*, "A methodology for bottom-up modelling of energy transitions in the industry sector: The FORECAST model," *Energy Strategy Rev.*, vol. 22, pp. 237–254, Nov. 2018.

Songmin Yu received the Ph.D. degree in management science from the Chinese Academy of Sciences, Beijing, China, in 2018. His Ph.D. dissertation is "A Study on the Carbon Emission Trading Scheme: An Agent-based Approach."

He is currently a Senior Researcher with the Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany. He works in the field of energy and climate policy with a special focus on the residential and tertiary sectors. His research interests include energy transition and climate policy, energy system modeling, and agent-based modeling. Before officially joining Fraunhofer ISI in May 2020, he worked at ISI as a Visiting Scholar for 18 months, supported by the Sino-German (CSC-DAAD) Postdoc Scholarship Program.