Applied Energy 314 (2022) 118871

Contents lists available at ScienceDirect

Applied Energy

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

Mechanism design for data sharing: An electricity retail perspective

Bohong Wang [a], Qinglai Guo [a,*], Yang Yu [b]

[a] Department of Electrical Engineering, Tsinghua University, Beijing 100084, China
[b] Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China

HIGHLIGHTS

- Data products can reduce uncertainty in decision-making problems with uncertainty.
- Data revenues and costs are defined by uncertainty reduction and privacy exposure.
- Electricity-side and data-side transactions are linked with data sharing framework.
- The profit allocation mechanism ensures the profitability of electricity retailer.

ARTICLE INFO

Keywords:
Data sharing
Uncertainty
Electricity retail
Profit allocation

ABSTRACT

Information incompletion always forces market participants to make decisions under uncertainty in energy transactions, while obtaining related data is a feasible way for them to reduce uncertainty and gain profits. However, the application of data transactions is not yet mature. To extend the application range of data transactions and concretize the data transaction model, a novel framework of electricity-side and data-side transactions linked with data sharing is proposed from the electricity retail perspective in this paper. The necessity and processes of data sharing between the electricity retailer and data suppliers are elaborately illustrated in the framework. Data revenues and data costs are analyzed according to uncertainty reduction and information provision. Considering the widely used two-settlement system of electricity markets, data revenues and data costs can be expressed in closed forms and their differences are net profits that are regarded to drive the data flow. Furthermore, an ex-post profit allocation mechanism is matched to appropriately allocate the net profits between the electricity retailer and data suppliers in the data sharing model. By comparison with the Shapley value method, the mechanism is less time-consuming and will ensure the profitability of the electricity retailer. Finally, a practical case with real data is employed to realize the results proposed in the theoretical analysis, and the feasibility of the data sharing model and profit allocation mechanism is validated.

1. Introduction

Information incompletion will lead to ambiguity and uncertainty in the decision-making process of stakeholders [1], which is common in energy systems, especially when renewable energies are involved [2,3]. For instance, on the source-side, renewable energy generation cannot be forecasted accurately, thus the lack of information generates risks to the decision-making of renewable energy producers when they trade electricity in the wholesale market. Similarly, on the load-side, electricity retailers usually do not fully know the actual demand of end users in the electricity retail market, thus the lack of information generates risks to the decision-making of electricity retailers when they purchase electricity from the wholesale market. The uncertainty will probably cause an electricity imbalance between day-ahead (DA) electricity bids and real-time (RT) electricity consumption, consequently resulting in economic losses [4].

From the perspective of electricity retail, the current literature proposes two approaches to manage the electricity retailer's risks caused by the information incompletion: pricing design and capturing the real preference of users. Generally, designing the retail price mechanism is an effective way to regulate end users' electricity consumption behavior for uncertainty reduction [5,6] or motivate end users to participate the demand response [7,8]. In existing studies, electricity retail is often modeled as a Stackelberg game [9,10], where electricity retailers take price elasticity and users' reaction to prices into account to design

* Corresponding author.
E-mail address: guoqinglai@tsinghua.edu.cn (Q. Guo).

https://doi.org/10.1016/j.apenergy.2022.118871
Received 17 November 2021; Received in revised form 31 January 2022; Accepted 26 February 2022
Available online 21 March 2022
0306-2619/© 2022 Elsevier Ltd. All rights reserved.

Nomenclature

Indices

i, j	Index of data suppliers
t	Index of time periods

Parameters

X_{min}, X_{max}	Minimum and maximum possible value of electricity demand (kWh)
σ	Standard deviation of RT electricity demand

Sets

$\mathscr{I}$	Set of all data suppliers
$\mathscr{I}_o$	Set of active data suppliers who can share data with the electricity retailer

Variables

π_t	Day-ahead (DA) market price in time period t ($/kWh)
p_t	Retail electricity price in time period t ($/kWh)
λ_t^+	Positive imbalance real-time (RT) price of electricity in time period t ($/kWh)
λ_t^-	Negative imbalance RT price of electricity in time period t ($/kWh)
Π_t	Total profits of the electricity retailer in electricity transactions in time period t ($)
$\widetilde{X}_t$	DA electricity offer amount of the electricity retailer in time period t (kWh)
X_t	RT electricity random variable of the electricity retailer in time period t (kWh)
x_t	RT electricity demand amount of the electricity retailer in time period t (kWh)
θ_t	Probability of RT electricity demand of the electricity retailer in time period t
d	Purchase price of the raw data ($/kWh)
β, β_i	Privacy sensitivity of end users related to the data supplier or the ith data supplier ($/kWh)
V_σ	Data value rate (marginal revenue) in electricity transactions for the electricity retailer ($/kWh)
$V_{\sigma i,eq}$	Equivalent data value rate in electricity transactions for the ith data supplier ($/kWh)
C_σ	Marginal cost in data acquisition for data suppliers ($/kWh)
σ_e	Standard deviation of overall load forecast error
σ_i	Standard deviation of load forecast error for ith data supplier
$\Delta\sigma, \Delta\sigma_i$	Reduction of the standard deviation of load forecast error for the data supplier or the ith data supplier
c_i	Cost performance of the ith data supplier in uncertainty reduction (($/kWh)$^{-0.5}$)
CS_1, CS_N	Consumers' surplus in data sharing for one electricity retailer and one data supplier or N data suppliers
t_o	Maximum number of the active data supplier with minimum cost performance
$CS_{i,eq}$	Profits of the ith data supplier
CS_{er}^d	Profits of the electricity retailer

Functions

$f_{X_t}(x_t)$	Probability density function (PDF) of RT electricity demand
$F_{X_t}(x_t)$	Cumulative distribution function (CDF) of RT electricity demand
$F_{X_t}^{-1}(\theta_t)$	Inverse function of CDF of RT electricity demand

optimal retail prices and even customized retail prices for different users. However, users' electricity consumption behavior is always affected by multiple intrinsic or cognitive factors, which are not merely related to retailer prices, e.g., ref. [11] found that users' perceptions of electricity bills can result in their behavior's deviation from the expectations in the retail price design. Therefore, the application of microeconomic principles and the Stackelberg game has encountered a bottleneck to some extent. To overcome this dilemma, some recent studies have carried out more realistic modeling of users' behavior and psychological analyses to refine the pricing mechanism [12,13], but the generality and accuracy of the analyses still remain to be seen.

We argue there exists another approach of managing the electricity retailer's risks due to the lack of end users' information, which is developing data transactions. Inspired by the assertion of Shannon in information theory that information eliminates uncertainty [14], obtaining user-related data is considered an essential way to break information barriers and reduce information incompletion and uncertainty in this paper. Note that it is not difficult for the retailer to obtain load data. In the current market system, electricity retailers usually can install smart meters for the end users, thus they will have access to the smart meter data, which can help them monitor the end users' electricity consumption. However, the smart meter data, or more precisely, the end users' historical electricity consumption data are not our target for data acquisition or data sharing. In our conception, end users' private data, such as their daily routines and travel plans [15], may contain the end users' future electricity consumption information, can help electricity retailers improve load forecast precision, avoid unnecessary deviation penalty, and thus guarantee revenues. For example, different lengths of travel in a day will affect an electric vehicle owner's electric vehicle charging load at night; the completion of daily work plans will affect the electricity consumption duration of office workers. These unstructured data cannot be obtained from smart meters or other instruments, but they will essentially affect end users' electricity consumption. However, in practice, collecting and analyzing a large quantity of user data is difficult for electricity retailers, especially when the users are usually not fixed [9]. Therefore, data suppliers are needed as intermediaries to address the inaccessibility of users' information. They serve as data brokers who are responsible for the search and collection of data [16] and the integration and sale of data products. Electricity retailers should obtain their required users' corresponding data products through data transactions with data suppliers to assist their decision-making.

In recent years, with the development of information technologies, the implementation of data transactions has made great progress in industry application and scientific research and will probably maintain the trend of rapid growth. Currently, the total amount of data in the world generated per day is approximately 2.5 quintillion bytes [17], and the large volume of data has also spawned several big data trading platforms worldwide, such as Factual, Infochimps, Microsoft Azure, and Fujitsu [18]. A noted data platform Statista forecasts that the annual revenue from the global big data and business analytics market is expected to reach $274.3 billion by 2022 [19].

In existing studies, data transactions are quite hot in the fields of the Internet of Things (IoT) and computer science but are rare in electricity-related fields.

Studies in the field of IoT mainly focus on cost-benefit analysis and business model design based on game theory or auction theory. A competitive data trading model is proposed in [20]. Four levels of stakeholders, including service consumers, a data service provider, data brokers, and data providers, are involved. In addition, two noncooperative competition games and two optimization problems are constructed to depict the behavior of stakeholders. Optimal contract-based mechanisms are designed in [21] to maximize data sellers' utility in the online

data trading market where only data sellers and buyers are considered. The dominant data sellers should balance the tradeoff between data sale profits and data privacy costs. A mixed sales mechanism with separately or bundled selling datasets for data vendors is introduced in [22]. Because the different data sources of data vendors can affect the market prices of data sets, a monopoly data market is analyzed. The big data market model composed of data vendors, data service providers, and service customers is adopted in [23]. The optimal raw data size and optimal data price for the data service provider are given based on the Bayesian digital goods auction.

Studies in the field of computer science concentrate more on algorithm design to elaborate the relationships between data and economic benefits from diverse perspectives in information economics. The problem of purchasing an unbiased estimator for a data demander is considered in [24]. Given the necessary variance level, a data analyst can use a combinatorial optimization problem to find an optimal convex combination of variances with the lowest costs. Similarly, the markets for private data are studied in [25] through differential privacy. In different auction mechanisms, different decision goals are given for data analysts. A real-time match mechanism for buying and selling the training data of machine learning tasks was proposed in [26]. The mechanism has the functions of truthful bidding, price updating, and revenue settlement and can be adapted to multiple commercial settings. Another highlight in [26] is a "robustness to replication" version of the Shapley algorithm to guarantee fairness among data providers.

Studies in electricity-related fields are few but insightful. A modified algorithm solution for data markets in the scenario of electricity markets for trading renewable energy sources is proposed in [27] referring to [26]. An information valuation paradigm is shown in [28], where the ex-ante valuation of photovoltaic power forecasting-related information is employed based on Shannon entropy and the non-noise ratio.

However, few studies have considered the data markets related to electricity markets and the transaction model of data sharing between data demanders and data suppliers. Furthermore, as small-scale interactions of data with specific purposes, data markets based on electricity markets are more suitable and feasible for sharing data than large-scale data markets. The data transaction model for electricity retailers and its corresponding profit allocation mechanism are preliminarily proposed in our previous work [29]. Then, some literature [30] has discussed the similar topic. Based on our previous work [15,29], a spot market framework of data sharing based on electricity markets is comprehensively constructed, and an ex-post profit allocation mechanism that matches the data sharing and considers the contribution of both electricity-side and data-side participants is developed in this paper. To the best of our knowledge, data sharing and its corresponding profit allocation based on the electricity transactions on the load-side have not been proposed in previous literature. In addition, the data sharing model can be extended to other similar goods markets. Different from the frameworks proposed in [26] and [31], which mainly focus on the regret of the electricity retailer, our framework for data sharing emphasizes the full cooperation between the electricity retailer and data suppliers and achieve a win–win situation between electricity-side and data-side participants.

The main contributions in this paper are as follows:

a) A complete framework of electricity-side and data-side transactions linked with data sharing is constructed from the perspective of electricity retail, where the electricity retailer and data suppliers are the main market participants. Based on the scenario of electricity retail, the operational processes and monetary flow analysis of data transactions are more specific;
b) An ex-post profit allocation mechanism for data sharing is creatively developed. In this mechanism, the data suppliers will be allocated profits first based on their uncertainty reduction and privacy sensitivity, while the residual profits will be allocated to the electricity retailer. In addition, the mechanism has the advantage of calculation efficiency and ensures profit maximization for the electricity retailer and data suppliers.

The remainder of this paper is organized as follows. In Section II, the framework of electricity-side and data-side transactions is constructed, and data revenues and data costs are analytically defined. Next, consumers' surplus maximization is solved, and a novel profit allocation mechanism between the electricity retailer and data suppliers is described in Section III. In Section IV, a practical case with real data is studied to demonstrate the theoretical results. Finally, conclusions are drawn and future work is briefly introduced in Section V.

2. Electricity-Side and Data-Side transactions

Here, we consider electricity-side and data-side transactions, where data-side transactions are designed to enable the electricity retailers to obtain the end users' information. The entire framework of the coupled transactions is shown in Fig. 1, and their operational processes are shown in Fig. 2. Data products are obtained in data-side transactions by data suppliers and utilized in electricity-side transactions by the electricity retailer; therefore, the net profits generated from transactions will be allocated between them. Detailed descriptions are as follows.

A. Framework

On the electricity-side, the electricity retailer participates in electricity wholesale and retail to earn profits. The wholesale market is a two-settlement market system, which includes a day-ahead (DA) market and a real-time (RT) balancing market. Due to the load demand uncertainty of end users, the inaccuracy of the load forecast is usually inevitable. Hence, the electricity retailer will probably suffer imbalance costs for electricity deviation [32,33]. To optimize his profits, the electricity retailer needs to avoid the risks that the load demand uncertainty brings to him as much as possible.

On the data-side, data suppliers glean private data about electricity consumption information from end users and develop mature data products for more accurate load forecasts through data processing, where the specific processes of data processing and load forecasting are well established in previous literature and will be briefly described in this paper as they are non-essential points. Data suppliers pay the privacy costs to satisfy and motivate end users, and the data products that reveal end users' future load demand may bring potential benefits to data suppliers. Besides, it is noted that different data suppliers collect raw data from different groups of end users, rather than the same group of end users, and their load forecasting data services are combined to form more accurate overall load forecasting results for the electricity retailer, thus the data suppliers can be regarded as the representatives of numerous data sources.

To match the demand of the electricity retailer in uncertainty reduction and the data products of data suppliers for uncertainty reduction, data sharing that links the electricity side and data side is proposed. Unlike [27], the "shared data" in this paper is more like data products or data services, which are similar to the "services" mentioned in [20] and [23]. Raw data are left on the data side, while only data products with typical data features (uncertainty reduction) are shared with the electricity side. Then, the electricity retailer and data suppliers will share net profits brought by data products. Data sharing and profit allocation, which are highlighted in the red box in Fig. 1, are the key content of this paper.

Based on the monetary flows, we focus on the revenues that the end users' private data will bring finally and the costs that the end users' private data will be spent initially to maximize the profits from data sharing and highlight the value-added of raw data in the flow from end users via data suppliers to the electricity retailer. The electricity retailer and data suppliers will absorb the added value and allocate it. Alternatively, end users can gain extra benefits from acceptable information

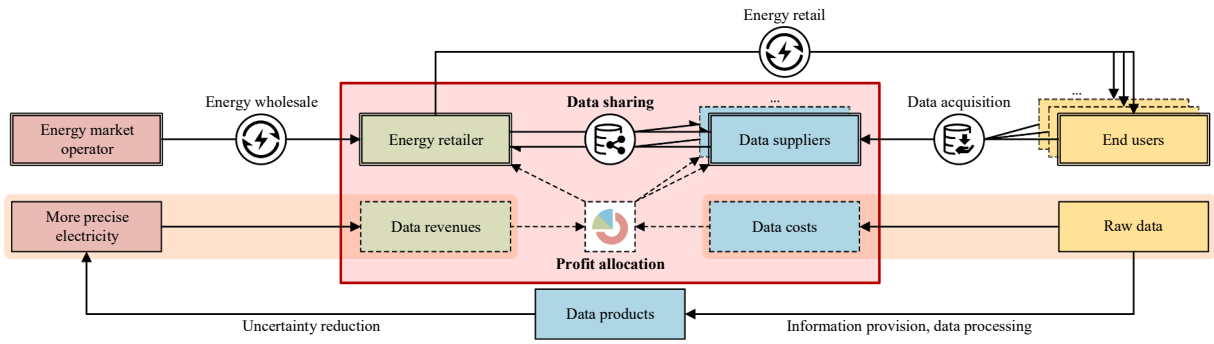

Fig. 1. Transaction framework of data acquisition, electricity transactions, and data sharing.

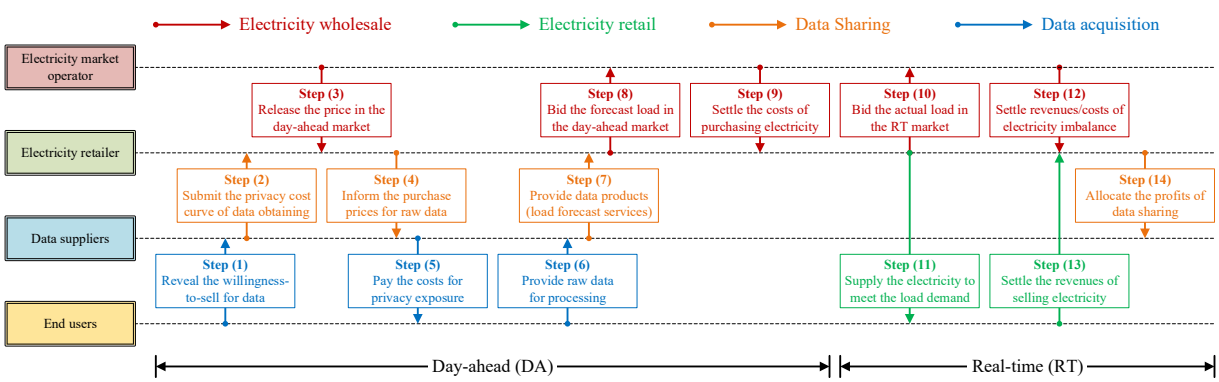

Fig. 2. Operational processes between the market operator, the electricity retailer, data suppliers, and end users in electricity transactions and data sharing.

provision. Therefore, data sharing makes electricity-side and data-side transactions complement each other and further achieves information asymmetry elimination and profit enhancement.

B. Electricity Transactions

1) Operational Processes

The operational processes of electricity transactions include Steps (3), (8), and (9) in the DA stage and Steps (10), (11), (12), and (13) in the RT stage, as shown in Fig. 2. In electricity transactions, the electricity retailer should purchase electricity from the wholesale market and retail it to end users.

In the DA stage, the electricity retailer submits his optimal bids to the market operator according to the DA price and his load forecast, then the electricity retailer should pay the costs for purchasing electricity; in the RT balancing market, the electricity retailer needs to adjust his bids, that is, the electricity retailer should sell electricity excesses at lower prices or buy electricity deficits at higher prices to meet users' requirements for electricity consumption.

After purchasing electricity in the wholesale market, the electricity retailer will directly sell it to end users. Finally, in profit settlement, the difference between retail revenues and wholesale costs will be the profits for the electricity retailer.

2) Data Revenues

The economic benefits gained in electricity transactions caused by uncertainty reduction are termed *data revenues*, which can be generally interpreted as the utility enhancement brought by data for data product demanders represented by the electricity retailer, namely, the pragmatic revenues of data.

Under the two-settlement electricity market, the profits of the electricity retailer in time period t are expressed as [15]

$$\Pi_t = \begin{cases} p_t x_t - \pi_t \widetilde{X}_t + \lambda_t^+ \left(\widetilde{X}_t - x_t\right), X_{min} \leqslant x_t < \widetilde{X}_t \\ p_t x_t - \pi_t \widetilde{X}_t - \lambda_t^- \left(x_t - \widetilde{X}_t\right), \widetilde{X}_t \leqslant x_t \leqslant X_{max} \end{cases} \tag{1}$$

We consider a fully competitive data market, thus, the electricity retailer is regarded as a price taker and cannot affect electricity market prices through the implementation of strategic offers [34]. The optimization model for a risk-neutral electricity retailer is

$$\max_{\widetilde{X}_t} \mathbb{E}\left[\Pi_t\left(\pi_t, \lambda_t^+, \lambda_t^-\right)\right] = \int_{X_{min}}^{X_{max}} \Pi_t \cdot f_{X_t}(x_t) dx_t$$

$$= \int_{X_{min}}^{\widetilde{X}_t} \left[p_t x_t - \pi_t \widetilde{X}_t + \lambda_t^+ \left(\widetilde{X}_t - x_t\right)\right] f_{X_t}(x_t) dx_t + \int_{\widetilde{X}_t}^{X_{max}} \left[p_t x_t - \pi_t \widetilde{X}_t - \lambda_t^- \left(x_t - \widetilde{X}_t\right)\right] f_{X_t}(x_t) dx_t \tag{2}$$

$$s.t. \qquad X_{min} \leqslant \widetilde{X}_t \leqslant X_{max} \tag{3}$$

Based on stochastic inventory theory and the newsvendor model [35], the optimal DA electricity bid and the corresponding optimal profits for the electricity retailer will be

$$\widetilde{X}_t^* = F_{X_t}^{-1}\left(\frac{\lambda_t^- - \pi_t}{\lambda_t^- - \lambda_t^+}\right), \lambda_t^+ < \pi_t < \lambda_t^- \tag{4}$$

$$\mathbb{E}\Pi_t^* = (p_t - \pi_t)\mathbb{E}X_t - V_\sigma\left(\pi_t, \lambda_t^-, \lambda_t^+\right)\cdot\sigma \tag{5}$$

where it is proven that the data value rate $V_\sigma(\pi_t, \lambda_t^-, \lambda_t^+)$ is a constant when electricity market prices $(\pi_t, \lambda_t^-, \lambda_t^+)$ and the distribution of X_t are given [15]. Therefore, (5) is an affine function to σ, thus, unit uncertainty reduction corresponds to the same expected profit enhancement.

The optimal profits $\mathbb{E}\Pi_t^*$ can be considered as the difference between expected no-risk revenues $(p_t - \pi_t)\mathbb{E}X_t$ and expected risk costs $V_\sigma(\pi_t, \lambda_t^-, \lambda_t^+)\cdot\sigma$. The expected no-risk revenues mainly depend on expected users' load demand, while the expected risk costs are related to the load demand uncertainty level represented by the standard deviation (SD) σ. In addition, although RT market prices are also uncertain, the expectations of λ_t^- and λ_t^+ in the past few time periods can replace them in (4)-(5) without affecting the results [36].

C. Data Acquisition

1) Operational Processes

The operational processes of obtaining data include Steps (1), (5), and (6) in the DA stage, as shown in Fig. 2. On the one hand, the day ahead end users' data are with a higher value for sharing, that is, data suppliers would be more willing to pay for obtaining day ahead end users' data; on the other hand, the modeling of the DA electricity market is easier for quantitative analysis. Therefore, the data acquisition is arranged in the DA market instead of long-term markets. In the process of data acquisition, data suppliers gather raw data according to the equilibrium price from end users.

In the beginning, data suppliers need to broadly investigate the willingness-to-sell of end users for their data and update it every other time period. Data suppliers also should develop a privacy cost curve of raw data and submit it to the electricity retailer. After the electricity retailer is informed about the privacy cost curves and makes the decision of profit maximization, data suppliers will purchase private raw data from end users at specific prices. In addition, the main processes of obtaining data are in the DA market stage to provide necessary assistance for the decision-making of the electricity retailer.

2) Data Costs

The economic costs for data suppliers in the process of data acquisition due to users' information provision are termed *data costs*, which can be generally interpreted as costs that need to be paid for the uncertainty reduction of the semantic form of data for data suppliers because information provision is regarded as the certainty of data that the end users transfer to the data supplier from the perspective of uncertainty in this paper.

Based on the results of laboratory experiments in [37], the willingness of users for information provision related to price signals is modeled in [38]. Considering that we need to use the unit prices of data for economic analysis, the equation of the willingness to sell can be modified as (6):

$$\Delta\sigma = \sigma_e\left[1 - exp\left(-\frac{d\Delta\sigma}{\beta}\right)\right] \tag{6}$$

where β is the privacy sensitivity to measure the users' unwillingness to information provision. If the user is more sensitive to information provision, the value of β will be larger, and the costs for others to obtain his data will be correspondingly higher.

Note that the economic incentives for providing the information of accurate planned electricity use are often a bonus for end users, instead of a necessity, and they are much lower than electricity bills. Therefore, end users are modeled as passive participants, who will be satisfied once they obtain the benefits corresponding to their information provision but have no intention to optimize the amount of information provision to maximize their revenues from it.

Eq. (6) implies that the unit expenditure of data suppliers will increase as they have achieved more uncertainty reduction, that is, the same unit data is much less helpful for data suppliers when they have achieved more uncertainty reduction. Besides, the key point in this paper is to elaborate the data value and design a reasonable mechanism for data sharing to make these private data work better, rather than to establish strict data sharing market rules, therefore, the effect of data is uniformly defined as the uncertainty reduction as Eq. (6) regardless of their differences in granularity, types or other characteristics.

3. Data sharing and its profit allocation mechanism

The operational processes of data sharing include Steps (2), (4), and (7) in the DA stage and Step (14) in the RT stage, as shown in Fig. 2. Note that data sharing is regarded to be used to serve electricity transactions. In this paper, the demand for uncertainty reduction for the electricity retailer gives birth to his demand for end users' data, and data suppliers appear to meet the demand of the electricity retailer because they can get allocated profits which is somewhat similar to the commission only when they share valuable end users' data with the electricity retailer. Therefore, the data sharing is dominated by the electricity retailer, and he is responsible for data pricing and profit allocation based on the market-clearing results in electricity transactions to guarantee enough profits for his community of economic interests with data suppliers. Besides, data revenues and data costs, as external economic characteristics, are emphasized; the possible costs of data suppliers in data processing, as an internal economic characteristic, are neglected.

A. Supply-Demand Curve

The initial SD of the distribution of the electricity retailer's load forecast error is denoted as σ_e, and the reduction of the SD, as a practical reflection of reduced uncertainty, is denoted as $\Delta\sigma$.

Differentiating (3), the marginal revenue of the data can be obtained in (7), namely, the data value rate V_σ. Because V_σ remains constant, the demand curve remains horizontal in the semi-open closed interval [0, σ_e) and drops to 0 at σ_e.

$$MR = V_\sigma = \left.\frac{d\mathbb{E}\Pi_t^*}{d\sigma}\right|_{\Delta\sigma} = V_\sigma\left(\pi_t, \lambda_t^-, \lambda_t^+\right) \tag{7}$$

Differentiating (6), the marginal cost of data C_σ can be obtained in (8):

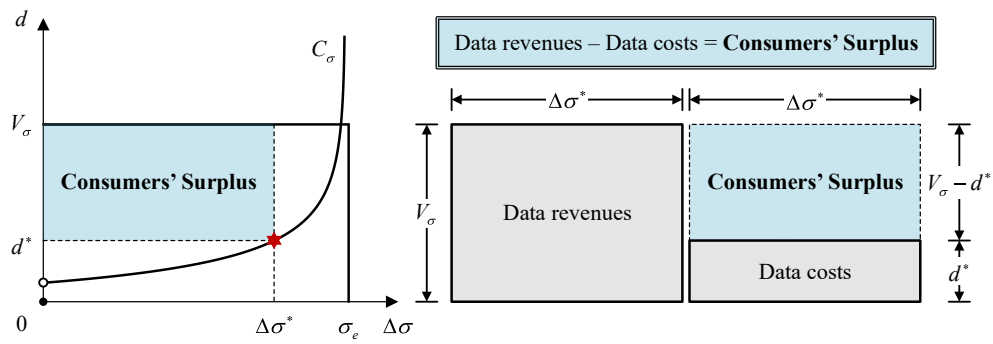

Fig. 3. Supply-demand curve of the uncertainty reduction and the unit price of data. (The dark red hexagonal star corresponds to the optimal uncertainty reduction and the optimal unit price of data for the maximum consumers' surplus.) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$MC = C_\sigma = d = \begin{cases} 0, & \Delta\sigma = 0 \\ \dfrac{\beta}{\Delta\sigma} ln \dfrac{\sigma_e}{\sigma_e - \Delta\sigma}, & 0 < \Delta\sigma < \sigma_e \end{cases} \tag{8}$$

Using the supply–demand analysis in microeconomics, the supply–demand curve of data can be drawn as in Fig. 3, where the axis of "quantity" is the uncertainty reduction $\Delta\sigma$, and the axis of "price" is the unit price of data d. In addition, $\Delta\sigma^*$ and d^* are the optimal uncertainty reduction and the optimal unit price of data corresponding to the maximum consumer surplus.

B. Consumers' Surplus Maximization

1) One Electricity Retailer and One Data Supplier

Considering that end users are passive participants, the objective of the electricity retailer and data suppliers in data sharing is to maximize the consumers' surplus regardless of end users' decision-making. To elaborate the consumers' surplus maximization, the optimization problem for one electricity retailer and one data supplier can be intuitively modeled as

$$\max_{\Delta\sigma} CS_1 = (V_\sigma - d)\Delta\sigma \tag{9}$$

$$s.t. \quad (6)$$

$$0 \leqslant \Delta\sigma < \sigma_e \tag{10}$$

Constraint (10) is derived from (8) and indicates the range of the uncertainty reduction.

The function of consumers' surplus to $\Delta\sigma$ is

$$CS_1 = V_\sigma \Delta\sigma - \beta ln \frac{\sigma_e}{\sigma_e - \Delta\sigma} \tag{11}$$

According to the first-order optimality condition, the optimal uncertainty reduction and the maximum consumer surplus are

$$\Delta\sigma^* = \sigma_e - \frac{\beta}{V_\sigma} \tag{12}$$

$$CS_1^* = V_\sigma \left(\sigma_e - \frac{\beta}{V_\sigma}\right) - \beta ln \frac{V_\sigma \sigma_e}{\beta} \tag{13}$$

The first term in (13) is the *data revenues* corresponding to the data valuation in Section II-B, and the second term is the *data costs* corresponding to the willingness to sell in Section II-C. The maximum consumer surplus merely depends on β and will decrease as β increases because uncertainty reduction is more difficult when β is larger.

In addition, note that the necessary and sufficient condition for the optimization problem to have a nonzero feasible solution is

$$\lim_{\Delta\sigma \to 0^+} d = \frac{\beta}{\sigma_e} < V_\sigma \tag{14}$$

2) One Electricity Retailer and N Data Suppliers

Based on the results in the simplest scenario, the number of data suppliers is extended to N to satisfy more general scenarios. Because the maximum consumer surplus is the profit enhancement brought by the data for the electricity retailer and N data suppliers, it should be well allocated between them.

To avoid unnecessary misunderstanding, the profit allocation mechanism for the electricity retailer and N data suppliers in this paper is based on the two assumptions below.

Assumption 1. *The electricity retailer and N data suppliers have symmetrical information before the data are obtained and shared; that is, the load forecast accuracy of the electricity retailer and N data suppliers are based on the same public information environment.*

Assumption 2. *The load forecast errors for different end-user groups' electricity consumption are uncorrelated, where different end-user groups are*

divided by data suppliers.

Therefore, the SDs of the load forecast errors of the electricity retailer and N data suppliers can be expressed according to the special case of the Bienaymé formula [24] as follows:

$$\sigma_e^2 = \sum_{i\in\mathscr{I}} \sigma_i^2 \tag{15}$$

where $\mathscr{I} = \{1, 2, \cdots, \mathrm{N}\}$ indicates the set of all data suppliers.

In the set $\mathscr{I}$, the serial numbers of data suppliers are arranged in descending order of c_i for the convenience of subsequent expressions, namely,

$$c_i \geqslant c_j (\forall i, j \in \mathscr{I}, \mathrm{i} < \mathrm{j}) \tag{16}$$

where c_i is defined in (17). σ_i denotes the uncertainty level, and β_i denotes the difficulty in uncertainty reduction. Hence, c_i can be understood as the cost performance in uncertainty reduction; that is, obtaining data products from the data supplier with higher c_i is more cost-effective for the electricity retailer.

$$c_i = \frac{\sigma_i}{\sqrt{\beta_i}}, \forall i \in \mathscr{I} \tag{17}$$

Then, the optimization problem for one electricity retailer and N data suppliers is constructed as

$$\max_{\Delta\sigma_i} CS_N = V_\sigma \Delta\sigma_e - \sum_{i\in\mathscr{I}} d_i \Delta\sigma_i \tag{18}$$

$$s.t. \qquad (15)$$

$$(\sigma_e - \Delta\sigma_e)^2 = \sum_{i\in\mathscr{I}} (\sigma_i - \Delta\sigma_i)^2 \tag{19}$$

$$\Delta\sigma_i = \sigma_i \left[1 - exp\left(-\frac{d_i \Delta\sigma_i}{\beta_i}\right)\right], \forall i \in \mathscr{I} \tag{20}$$

$$0 \leqslant \Delta\sigma_i < \sigma_i, \forall i \in \mathscr{I} \tag{21}$$

Constraint (19) indicates that the SDs of load forecast errors of the electricity retailer and N data suppliers after data sharing still satisfy the Bienaymé formula; Constraint (20) indicates that the characteristic of willingness to sell is suitable for all data suppliers; Constraint (21) is similar to (10).

The function of consumers' surplus to $\Delta\sigma_i$ is

$$CS_N = V_\sigma \left(\sigma_e - \sqrt{\sum_{i\in\mathscr{I}} (\sigma_i - \Delta\sigma_i)^2}\right) - \sum_{i\in\mathscr{I}} \beta_i ln \frac{\sigma_i}{\sigma_i - \Delta\sigma_i} \tag{22}$$

Considering that obtaining the load forecasting data from some data suppliers is not cost-effective according to the supply–demand curve in Fig. 3, which implies that not all data suppliers can successfully share data with the electricity retailer, thus data suppliers should be divided into active and inactive data suppliers based on whether their data are enough cost-effective for data sharing. Sets $\mathscr{I}_\mathrm{o}$ and $\mathscr{I}\backslash\mathscr{I}_\mathrm{o}$ are used to denote them. The electricity retailer will only share data products with active data suppliers and allocate profits to them. Mathematically, active data suppliers' $\Delta\sigma_i$ are nonzero, but inactive data suppliers' $\Delta\sigma_i$ are zero. However, because parameters in the transactions are time-variant, elements in Sets $\mathscr{I}_\mathrm{o}$ and $\mathscr{I}\backslash\mathscr{I}_\mathrm{o}$ are not fixed. According to the optimal conditions, set $\mathscr{I}_\mathrm{o}$ is expressed as:

$$\mathscr{I}_\mathrm{o} = \left\{ \mathrm{i} \in \mathscr{I} \middle| \mathrm{c_i} > \frac{1}{\sqrt{2}\,\mathrm{V}_\sigma} \sqrt{\sum_{\mathrm{j}=1}^{\mathrm{t_o}} \beta_\mathrm{j} + \sqrt{\left(\sum_{\mathrm{j}=1}^{\mathrm{t_o}} \beta_\mathrm{j}\right)^2 + 4\mathrm{V}_\sigma^2 \sum_{\mathrm{j}=\mathrm{t_o}+1}^{\mathrm{N}} \sigma_\mathrm{j}^2}} \right\} \tag{23}$$

where t_o is defined to guarantee that all data suppliers with the same minimum c_i in $\mathscr{I}_\mathrm{o}$ are involved, that is, all data suppliers with the least cost-effective data for active data suppliers are divided into the group of active data suppliers because data suppliers with the same c_i must be either all active or all inactive.

$$t_o = max\left(\underset{\mathrm{i}\in\mathscr{I}_\mathrm{o}}{argmin} c_i\right) \tag{24}$$

Therefore, the optimality conditions can be solved. To be similar to the expression form in (12), the optimal uncertainty reduction for each data supplier can be expressed as

$$\Delta\sigma_i^* = \begin{cases} \sigma_i - \dfrac{\beta_i}{V_{\sigma i,eq}}, i \in \mathscr{I}_\mathrm{o} \\ 0, \qquad\qquad\qquad\qquad i \in \mathscr{I}\backslash\mathscr{I}_\mathrm{o} \end{cases} \tag{25}$$

where $V_{\sigma i,eq}$ is considered the equivalent data value rate for each active data supplier and can be expressed as

$$V_{\sigma i,eq} = V_\sigma \frac{\sqrt{2\beta_i}}{\sqrt{\sum_{j\in\mathscr{I}_\mathrm{o}} \beta_j + \sqrt{\left(\sum_{j\in\mathscr{I}_\mathrm{o}} \beta_j\right)^2 + 4V_\sigma^2 \sum_{j\in\mathscr{I}\backslash\mathscr{I}_\mathrm{o}} \sigma_j^2}}}, \forall i \in \mathscr{I}_\mathrm{o} \tag{26}$$

The overall uncertainty reduction of the electricity retailer can be derived based on (19) and (25). Then, the maximum total consumers' surplus is

$$CS_N^* = V_\sigma\left(\sigma_e - \sqrt{\sum_{i\in\mathscr{I}_o}\frac{\beta_i^2}{V_{\sigma i,eq}^2} + \sum_{i\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_i^2}\right) - \sum_{i\in\mathscr{I}_o}\beta_i ln\frac{V_{\sigma i,eq}\sigma_i}{\beta_i} \tag{27}$$

Similar to (13), the first term in (27) is the *data revenues* and the second term is the *data costs*.

C. Profit Allocation Mechanism

For profit allocation, the electricity retailer will set the purchase price of data products with uncertainty reduction, and data suppliers will make decisions on obtaining data based on their end users' privacy sensitivity.

The purchase price of data products for the electricity retailer released to the ith active data supplier is $V_{\sigma i,eq}$; that is, the ith active data supplier will receive $V_{\sigma i,eq}$ as a reward when his data products can reduce the unit uncertainty of $\Delta\sigma_i$. Then, the optimization model for active data suppliers is similar to that mentioned in Section III-B. According to the results in (13), the maximum profits of the ith data supplier are shown in (28) and are positive:

$$CS_{i,eq}^* = V_{\sigma i,eq}\sigma_i - \beta_i - \beta_i ln\frac{V_{\sigma i,eq}\sigma_i}{\beta_i} > V_{\sigma i,eq}\sigma_i - \beta_i - \beta_i\left(\frac{V_{\sigma i,eq}\sigma_i}{\beta_i} - 1\right) = 0, \forall i \in \mathscr{I}_o \tag{28}$$

The maximum profits of data suppliers will increase as σ_i increases and decrease as β_i increases. In addition, among different active data suppliers, the electricity retailer should set equivalent purchase prices to motivate data suppliers to provide data products consistent with their privacy sensitivity and to achieve their maximum profits, which is similar to the constant micro-increment principle in economic dispatch. Therefore, a larger β_i corresponds to a larger $V_{\sigma i,eq}$ for active data suppliers; thus, the profits of data suppliers are approximately proportional to $\sigma_i\sqrt{\beta_i}$ considering (26) and (28).

Correspondingly, the profits of the electricity retailer are

$$CS_{er}^d = CS_N^* - \sum_{i\in\mathscr{I}_o} CS_{i,eq}^* = V_\sigma\left(\sigma_e - \sqrt{\sum_{i\in\mathscr{I}_o}\frac{\beta_i^2}{V_{\sigma i,eq}^2} + \sum_{i\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_i^2}\right) - \sum_{i\in\mathscr{I}_o}\left(V_{\sigma i,eq}\sigma_i - \beta_i\right) \tag{29}$$

In such a profit allocation mechanism, the two following unique properties should be emphasized.

Proposition 1. *The profits of the electricity retailer are nonnegative under the profit allocation mechanism.*

Proof: See Appendix A.

Proposition 2. *The electricity retailer and data suppliers both achieve their maximum profits in the proposed profit allocation mechanism.*

Proof: See Appendix B.

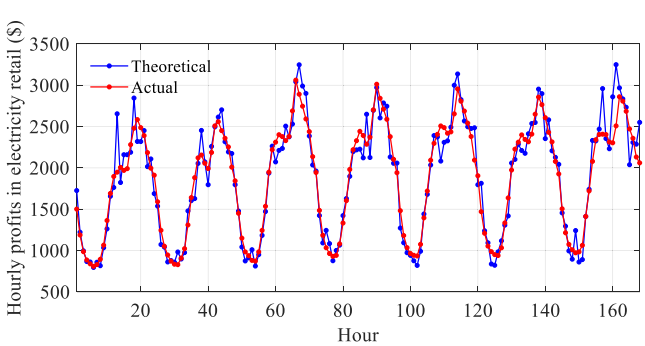

Fig. 4. Comparison of theoretical and actual hourly total profits for the electricity retailer in electricity wholesale and retail.

Table 1
Parameters of Four Data Suppliers.

Parameter	Data suppliers (DS)			
	#1	#2	#3	#4
Uncertainty level σ_i	0.04	0.06	0.05	0.02
Privacy sensitivity β_i	0.05	0.13	0.3617	0.05
Cost performance c_i	0.1789	0.1664	0.1213	0.0894

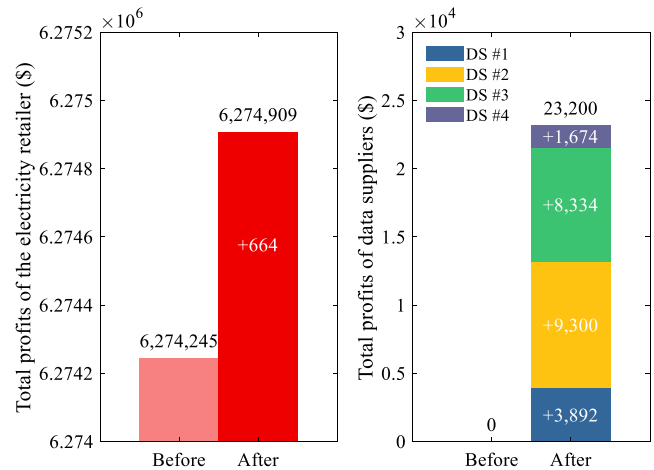

Fig. 5. Total profits of the electricity retailer and data suppliers before and after the data sharing.

4. Case study

Smart meter data from the Commission for Electricity Regulation (CER) based in Ireland serve as end users' electricity consumption data, and price data from the PJM electricity market in the US are served as the wholesale electricity market prices. For the consistency of data, the time period is determined to be one hour. In addition, the flat retail price is set to 0.2 $/kWh. λ_t^- and λ_t^+ are proportional to real-time market prices, and penalty rates for positive and negative electricity imbalance are both set to 0.5. Finally, the profit calculation programs are run on a personal computer with an Intel Core i7-8550U 1.80-GHz CPU and 16 GB of RAM.

A. Results of Data Sharing

1) Electricity Transactions and Profit Calculation

In our previous work [15], the estimation and calculation of data value rates were analyzed in detail. Although the Gaussian distribution is typically used in the distribution fitting for load forecast errors, the approximate data value rates of Gaussian distributions are not precise enough because its cumulative distribution function (CDF) cannot be expressed in closed form. Therefore, the logistic distribution, which is similar to the Gaussian distribution in shape, is used as the proxy to depict the distribution of load forecast errors.

The inverse function of the CDF of the logistic distribution is shown in (30). Several independent random variables that conform to the logistic distribution can be obtained by generating several independent random variables uniformly distributed on [0,1] and substituting them into (30) as the independent variables.

$$F_{X_t}^{-1}(\theta_t) = \mathbb{E}X_t + \frac{\sqrt{3}\sigma}{\pi} ln\frac{\theta_t}{1-\theta_t}, \theta_t \in (0,1) \tag{30}$$

Thus, the corresponding data value rate is [15]

$$V_{\sigma,t} = \frac{\sqrt{3}}{\pi}\left[\left(\lambda_t^- - \pi_t\right) ln\frac{\lambda_t^- - \lambda_t^+}{\lambda_t^- - \pi_t} + \left(\pi_t - \lambda_t^+\right) ln\frac{\lambda_t^- - \lambda_t^+}{\pi_t - \lambda_t^+}\right] \tag{31}$$

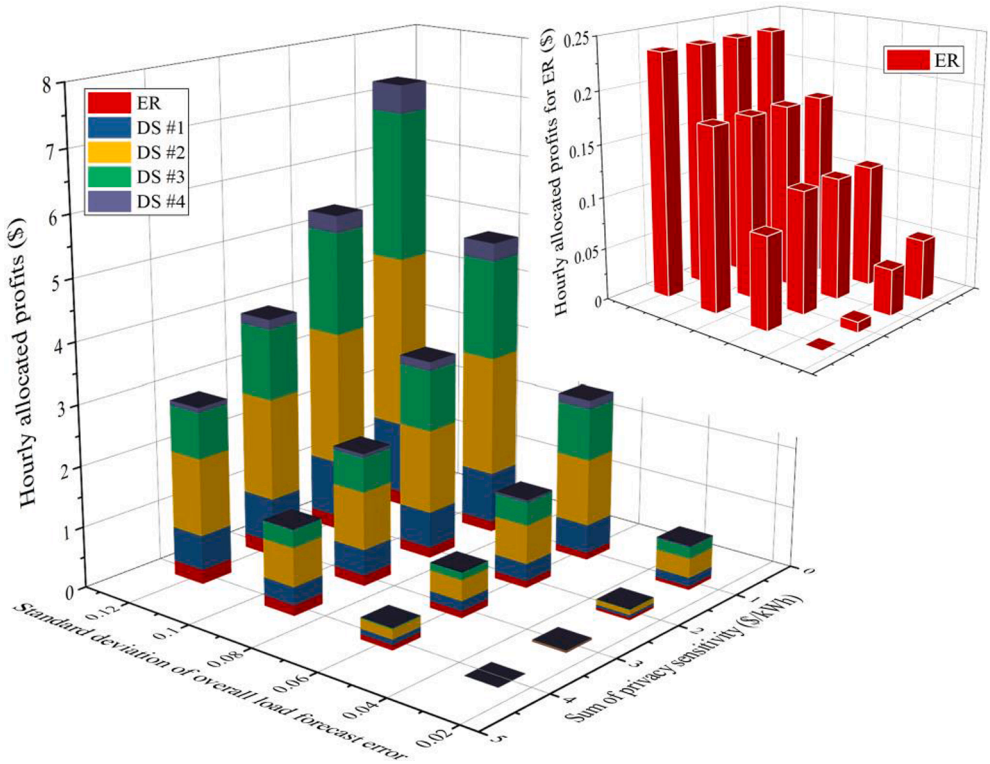

Fig. 6. Hourly allocated profits for the electricity retailer and data suppliers in the 13th hour in the 1st daty under differenct σ_e and $\Sigma\beta_i$.

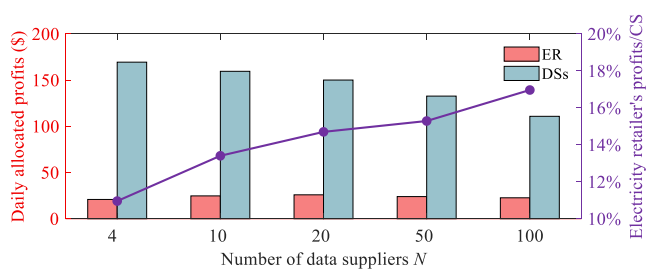

Fig. 7. Daily allocated profits for the electricity retailer and data suppliers in the 2nd day under different number of data suppliers.

σ_e is set to 0.09; then, the theoretical results for optimal expected profits are calculated by substituting (31) into (5), and their actual results are calculated based on randomly generated load forecast errors with σ_e as the SD and market price data. A week of hourly theoretical and actual total profits for the electricity retailer are compared in Fig. 4.

Fig. 4 shows that the theoretical total profits are very close to the actual total profits. Through further calculation, it is found that the error of total profits for the electricity retailer in half a year is merely 0.91%. Therefore, the theoretical results of data value are accurate enough to be the benchmark for profit allocation settlement in data sharing.

2) Data Sharing and Profit Allocation

In the data sharing, four data suppliers are set, and their corresponding parameters, including the SD of the load forecast errors σ_i, privacy sensitivity β_i, and cost performance in uncertainty reduction c_i, are shown in Table 1. In addition, the four data suppliers are arranged in descending order of c_i.

The profits of the electricity retailer and data supplier before and after data transactions are shown in Fig. 5. It can be found that they all gain additional profits from data sharing. Under the proposed profit allocation mechanism, the electricity retailer gains $664 and data

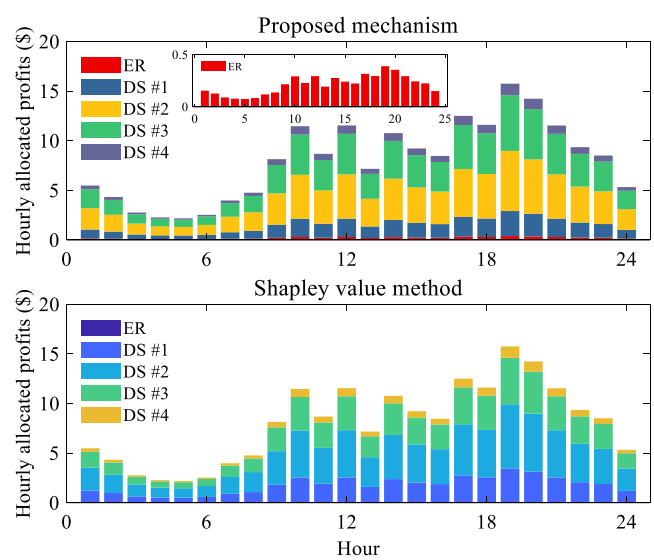

Fig. 8. Hourly allocated profits for the electricity retailer and data suppliers in the 49th day under the proposed mechanism and Shapley value method.

Table 2
Average hourly profit calculation time for the proposed mechanism and shapley value method.

Profit allocation method	Calculation time for N data suppliers (s)			
	$N = 4$	$N = 6$	$N = 8$	$N = 10$
Proposed mechanism	**0.0539**	**0.0674**	**0.0739**	**0.0779**
Shapley value method	0.7788	3.6185	15.7462	48.2543

suppliers #1∼#4 gain \$3,892, \$9,300, \$8,334, and \$1,674, respectively, in electricity transactions and data sharing for half a year. Therefore, the conjecture that the profits of data suppliers are approximately proportional to in Section III -C can be validated.

B. Sensitivity Analysis

1) Parameters of Data Suppliers

The SD of the overall load forecast error σ_e is set to 0.03, 0.06, 0.09, and 0.12, and the σ_i of each data supplier will change in the same proportion as σ_e. The sum of data suppliers' privacy sensitivity $\Sigma\beta_i$ is set to 1, 2, 3, and 4 \$/kWh, and the β_i of each data supplier will also change in the same proportion as $\Sigma\beta_i$.

Fig. 6 shows that the profits of data suppliers increase as σ_i increases and decrease as β_i increases. As interpreted in Section III -B, a larger σ_i indicates more uncertainty that needs to be reduced, and a larger β_i indicates that it is harder for the ith data supplier to obtain data and reduce the uncertainty. Thus, it is known that the profits that data suppliers receive will increase as the uncertainty that can be reduced increases and the difficulty of uncertainty reduction decreases.

Fig. 6 also shows that the trend of the electricity retailer's profits is similar to that of data suppliers, while the electricity retailer's profits will remain unchanged as β_i increases when σ_e is large enough. It can be found that the requirement of being an active data supplier in (24) is easier to achieve as σ_e increases. If all data suppliers are active data suppliers, the supreme of the electricity retailer's profits can be derived as (32) based on (29). Due to the setting that the β_i of each data supplier changes in the same proportion as $\Sigma\beta_i$, the electricity retailer's profits will reach and remain at its supreme when σ_i is fixed.

$$CS^{d}_{er,max} = V_\sigma\left(\sigma_e - \sum_{i\in\mathcal{J}}\sqrt{\frac{\beta_i}{\sum_{i\in\mathcal{J}}\beta_i}}\sigma_i\right) \tag{32}$$

2) Number of Data Suppliers

We set the number of data suppliers N equal to 4, 10, 20, 50, and 100 to perform the analysis. For the parameters of data suppliers, σ_e is set to 0.09, and σ_i of data suppliers is randomly generated subject to (15). In addition, the privacy sensitivity for each data supplier β_i is set to 0.01 \$/kWh to ensure fairness among data suppliers.

Fig. 7 shows the daily profits for the electricity retailer and data suppliers under different numbers of data suppliers. The profits of the electricity retailer do not change much, while the profits of data suppliers decrease as the number of data suppliers N increases. The equivalent difficulty for the electricity retailer to reduce the same uncertainty increases as the number of data suppliers N increases when σ_e and β_i are fixed. Therefore, the profits of data suppliers will decrease. For the electricity retailer, revenues through uncertainty reduction and profits allocated to data suppliers decrease synchronously; thus, his profits do not change significantly.

C. Comparison with Other Profit Allocation Methods

Shapley value is a commonly used profit allocation method and allocates profits based on the expected marginal contribution of each coalition member for fairness [39] because it gives every member the same chance of being in any position [40]. Note that the electricity retailer should always be a member of the coalition in data sharing; otherwise, the coalition composed of only data suppliers will not have any profit inflow.

Fig. 8 shows hourly profits for the electricity retailer and data suppliers under the proposed mechanism and Shapley value method. The electricity retailer can only gain profits under the proposed mechanism because his expected marginal contribution is not considered in the Shapley value method because he should be contained in each coalition. Data suppliers can gain similar profits under the two different methods, which shows the fairness of our proposed mechanism because the proposed mechanism can be regarded as an equilibrium converged to by the Shapley value method when data suppliers are price-takers.

In addition, because the allocated profits can be directly calculated by (28) and (29) instead of traversing each possible coalition, the average time for hourly profit calculation using the proposed mechanism is much less than that using the Shapley value method, as shown in Table 2. As the number of data suppliers N increases, the calculation time for the proposed mechanism increases slightly, while the calculation time for the Shapley value method increases rapidly and is unacceptable for the spot market when N is too large.

5. Conclusions and future work

This paper presents a complete framework of electricity-side and data-side transactions to propose a spot data sharing mode between an electricity retailer and data suppliers from the perspective of electricity retail. The electricity retailer's demand of more data for uncertainty reduction is considered, thus data suppliers who develop data products to meet his demand are introduced. The supply and demand of data products can be analyzed using analytical expressions based on the data valuation and the modeling of the end users' information provision. Through the solution of consumers' surplus maximization, an ex-post profit allocation mechanism is finally proposed to fairly allocate net profits in data sharing. The results in case studies validate that the data sharing and profit allocation mechanism can help the electricity retailer and data suppliers earn additional profits. In addition, the proposed mechanism is demonstrated to have advantages in the electricity retailer's profits and calculation time compared with the Shapley value method.

Future work related to data sharing will extend the model from load-side to source-side, incorporate the information about renewable energy forecasts, and conduct a more complex market design with multiple types of data demanders and suppliers. Moreover, an equivalent transformation of data and energy concerning the flexibility of integrated energy systems is another interesting derivative topic.

CRediT authorship contribution statement

Bohong Wang: Methodology, Software, Validation, Formal analysis, Visualization, Writing – original draft, Investigation, Data curation. **Qinglai Guo:** Supervision, Project administration, Funding acquisition, Writing – review & editing. **Yang Yu:** Methodology, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Key Research and Development Program of China 2020YFB2104500.

Appendix

A. Proof of Proposition 1

Considering (26), the first term in (29) is simplified through the construction of a complete square expression.

$$\sqrt{\sum_{i\in\mathscr{I}_o}\frac{\beta_i^2}{V_{\sigma i,eq}^2}+\sum_{i\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_i^2}=\frac{\sum_{j\in\mathscr{I}_o}\beta_j+\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{j\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_j^2}}{2V_\sigma} \tag{A1}$$

Thus, (29) can be otherwise expressed as

$$CS_{er}^d=V_\sigma\sigma_e-\frac{1}{2}\left(\sum_{j\in\mathscr{I}_o}\beta_j+\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{i\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_i^2}\right)-\sum_{i\in\mathscr{I}_o}\left(\frac{V_\sigma\sqrt{2\beta_i}\,\sigma_i}{\sqrt{\sum_{j\in\mathscr{I}_o}\beta_j+\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{j\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_j^2}}}-\beta_i\right) \tag{A2}$$

Then, (A2) can be further changed by the merger of similar items and the reduction of fractions to a common denominator.

$$\begin{aligned}CS_{er}^d&=V_\sigma\sigma_e-\frac{1}{2}\left(\sum_{j\in\mathscr{I}_o}\beta_j+\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{i\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_i^2}\right)-\sum_{i\in\mathscr{I}_o}\left(\frac{V_\sigma\sqrt{2\beta_i}\,\sigma_i}{\sqrt{\sum_{j\in\mathscr{I}_o}\beta_j+\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{j\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_j^2}}}-\beta_i\right)\\&=V_\sigma\sigma_e-V_\sigma\frac{\sum_{i\in\mathscr{I}_o}\sqrt{2\beta_i}\sigma_i}{\sqrt{\sum_{j\in\mathscr{I}_o}\beta_j+\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{j\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_j^2}}}-\frac{1}{2}\left(\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{i\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_i^2}-\sum_{j\in\mathscr{I}_o}\beta_j\right)\\&=V_\sigma\sigma_e-V_\sigma\frac{\sum_{i\in\mathscr{I}_o}\sqrt{2\beta_i}\sigma_i}{\sqrt{\sum_{j\in\mathscr{I}_o}\beta_j+\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{j\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_j^2}}}-V_\sigma\frac{\sqrt{\sum_{i\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_i^2}\sqrt{\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{i\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_i^2}-\sum_{j\in\mathscr{I}_o}\beta_j}}{\sqrt{\sum_{j\in\mathscr{I}_o}\beta_j+\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{j\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_j^2}}}\end{aligned} \tag{A3}$$

According to the Cauchy-Schwarz inequality, the numerator of the second term in (A3) can be scaled as

$$\sqrt{\sum_{i\in\mathscr{I}_o}2\beta_i}\sqrt{\sum_{i\in\mathscr{I}_o}\sigma_i^2}\geqslant\sum_{i\in\mathscr{I}_o}\sqrt{2\beta_i}\sigma_i \tag{A4}$$

The condition of equality in inequality (A.4) is

$$c_i=c_j,\forall i,j\in\mathscr{I}_o \tag{A5}$$

According to the Cauchy-Schwarz inequality, the second and the third terms in (A3) can be scaled as

$$\sigma_e=\sqrt{\sum_{i\in\mathscr{I}}\sigma_i^2}\geqslant\frac{\sqrt{\sum_{i\in\mathscr{I}_o}2\beta_i}\sqrt{\sum_{i\in\mathscr{I}_o}\sigma_i^2}}{\sqrt{\sum_{j\in\mathscr{I}_o}\beta_j+\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{j\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_j^2}}}+\frac{\sqrt{\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{i\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_i^2}-\sum_{j\in\mathscr{I}_o}\beta_j}\sqrt{\sum_{j\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_j^2}}{\sqrt{\sum_{j\in\mathscr{I}_o}\beta_j+\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{j\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_j^2}}} \tag{A6}$$

The condition of equality in inequality (A6) is

$$\frac{\sum_{i\in\mathscr{I}_o}\sigma_i^2}{2\sum_{i\in\mathscr{I}_o}\beta_i}=\frac{\sum_{j\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_j^2}{\sqrt{\left(\sum_{j\in\mathscr{I}_o}\beta_j\right)^2+4V_\sigma^2\sum_{i\in\mathscr{I}\setminus\mathscr{I}_o}\sigma_i^2}-\sum_{j\in\mathscr{I}_o}\beta_j} \tag{A7}$$

Merging (A5) and (A7), the condition that the profits of the electricity retailer equal zero are

$$c_i=\sqrt{\frac{\sigma_e}{V_\sigma}},\forall i\in\mathscr{I}_o \tag{A8}$$

Thus, the profits of the electricity retailer are mathematically nonnegative. Considering that the condition in (A8) is extremely strong, it is quite possible in practical scenarios that the profits of the electricity retailer are positive.

$$CS_{er}^{d} \geq V_{\sigma}\sigma_{e} - V_{\sigma}\sigma_{e} = 0 \tag{A9}$$

■

B. Proof of Proposition 2

Given the purchase prices of the electricity retailer for raw data, the consumer's surplus for the electricity retailer can be expressed as

$$CS_{er} = CS_{N} - \sum_{i\in\mathscr{I}_{o}} CS_{i,eq} = V_{\sigma}\left(\sigma_{e} - \sqrt{\sum_{i\in\mathscr{I}_{o}}(\sigma_{i}-\Delta\sigma_{i})^{2} + \sum_{i\in\mathscr{I}\setminus\mathscr{I}_{o}}\sigma_{i}^{2}}\right) - \sum_{i\in\mathscr{I}_{o}} V_{\sigma i,eq}\Delta\sigma_{i} \tag{B1}$$

Then, the partial derivatives of the consumers' surplus for the electricity retailer to the uncertainty reductions of each data supplier are

$$\begin{aligned}\frac{\partial CS_{er}}{\partial\Delta\sigma_{i}} &= \frac{\partial\left[V_{\sigma}\left(\sigma_{e} - \sqrt{\sum_{i\in\mathscr{I}_{o}}(\sigma_{i}-\Delta\sigma_{i})^{2} + \sum_{i\in\mathscr{I}\setminus\mathscr{I}_{o}}\sigma_{i}^{2}}\right) - \sum_{i\in\mathscr{I}_{o}} V_{\sigma i,eq}\Delta\sigma_{i}\right]}{\partial\Delta\sigma_{i}} \\ &= -V_{\sigma}\frac{\partial\sqrt{\sum_{i\in\mathscr{I}_{o}}(\sigma_{i}-\Delta\sigma_{i})^{2} + \sum_{i\in\mathscr{I}\setminus\mathscr{I}_{o}}\sigma_{i}^{2}}}{\partial\Delta\sigma_{i}} - V_{\sigma i,eq} \\ &= V_{\sigma}\frac{\sigma_{i}-\Delta\sigma_{i}}{\sqrt{\sum_{i\in\mathscr{I}_{o}}(\sigma_{i}-\Delta\sigma_{i})^{2} + \sum_{i\in\mathscr{I}\setminus\mathscr{I}_{o}}\sigma_{i}^{2}}} - V_{\sigma i,eq}, \forall i\in\mathscr{I}_{o}\end{aligned} \tag{B2a}$$

$$\frac{\partial CS_{er}}{\partial\Delta\sigma_{i}} = 0, \forall i \in \mathscr{I}\mathscr{I}_{o} \tag{B2b}$$

Using the optimal solution in (29), the partial derivatives of the consumers' surplus for the electricity retailer to the uncertainty reductions of each active data supplier equal zero.

$$\left.\frac{\partial CS_{er}}{\partial\Delta\sigma_{i}}\right|_{\Delta\sigma_{i}^{*}} = V_{\sigma}\frac{\sigma_{i}-\Delta\sigma_{i}^{*}}{\sqrt{\sum_{i\in\mathscr{I}_{o}}(\sigma_{i}-\Delta\sigma_{i}^{*})^{2} + \sum_{i\in\mathscr{I}\setminus\mathscr{I}_{o}}\sigma_{i}^{2}}} - V_{\sigma i,eq} = V_{\sigma}\frac{\frac{\beta_{i}}{V_{\sigma i,eq}}}{\sqrt{\sum_{i\in\mathscr{I}_{o}}\frac{\beta_{i}^{2}}{V_{\sigma i,eq}^{2}} + \sum_{i\in\mathscr{I}\setminus\mathscr{I}_{o}}\sigma_{i}^{2}}} - V_{\sigma i,eq} = 0, \forall i\in\mathscr{I}_{o} \tag{B3}$$

Therefore, the uncertainty reduction corresponding to each active data supplier's maximum profits also corresponds to the maximum profits of the electricity retailer and the maximum consumers' surplus in data sharing, namely,

$$\max_{\Delta\sigma_{i}} CS_{er} = CS_{er}^{d} \tag{B4}$$

References

[1] Laffont J-J, Martimort D. The Theory of Incentives: The Principal-Agent Model. Priceton, NJ, USA: Priceton University Press; 2002.

[2] González-Aparicio I, Zucker A. Impact of wind power uncertainty forecasting on the market integration of wind energy in Spain. Appl Energy Dec. 2015;159: 334–49. https://doi.org/10.1016/j.apenergy.2015.08.104.

[3] Exizidis L, Kazempour SJ, Pinson P, de Greve Z, Vallée F. Sharing wind power forecasts in electricity markets: A numerical analysis. Appl Energy Aug. 2016;176: 65–73. https://doi.org/10.1016/j.apenergy.2016.05.052.

[4] Bitar EY, Rajagopal R, Khargonekar PP, Poolla K, Varaiya P. Bringing Wind Energy to Market. IEEE Trans Power Syst Aug. 2012;27(3):1225–35. https://doi.org/10.1109/TPWRS.2012.2183395.

[5] Mamounakis I, Efthymiopoulos N, Vergados DJ, Tsaousoglou G, Makris P, Varvarigos EM. A pricing scheme for electric utility's participation in day-ahead and real-time flexibility energy markets. J Mod Power Syst Clean Energy Sep. 2019; 7(5):1294–306. https://doi.org/10.1007/s40565-019-0537-2.

[6] Liu Y, Zhang D, Gooi HB. Data-driven decision-making strategies for electricity retailers: A deep reinforcement learning approach. CSEE J. Power Energy Syst. Mar. 2021;7(2):358–67. https://doi.org/10.17775/CSEEJPES.2019.02510.

[7] Huang W, Zhang N, Kang C, Li M, Huo M. From demand response to integrated demand response: review and prospect of research and application. Prot Control Mod Power Syst May 2019;4(1):12. https://doi.org/10.1186/s41601-019-0126-4.

[8] Arun SL, Selvan MP. Smart residential energy management system for demand response in buildings with energy storage devices. Front Energy Dec. 2019;13(4): 715–30. https://doi.org/10.1007/s11708-018-0538-2.

[9] Feng C, Wang Y, Zheng K, Chen Q. Smart Meter Data-Driven Customizing Price Design for Retailers. IEEE Trans Smart Grid May 2020;11(3):2043–54. https://doi.org/10.1109/TSG.2019.2946341.

[10] Yang J, Zhao J, Luo F, Wen F, Dong ZY. Decision-Making for Electricity Retailers: A Brief Survey. IEEE Trans Smart Grid Sep. 2018;9(5):4140–53. https://doi.org/10.1109/TSG.2017.2651499.

[11] White LV, Sintov ND. Inaccurate consumer perceptions of monetary savings in a demand-side response programme predict programme acceptance. Nat Energy Dec. 2018;3(12):1101–8. https://doi.org/10.1038/s41560-018-0285-y.

[12] Good N. Using behavioural economic theory in modelling of demand response. Appl Energy Apr. 2019;239:107–16. https://doi.org/10.1016/j.apenergy.2019.01.158.

[13] Gao J, Ma Z, Yang Y, Gao F, Guo G, Lang Y. The Impact of Customers' Demand Response Behaviors on Power System With Renewable Energy Sources. IEEE Trans Sustain Energy Oct. 2020;11(4):2581–92. https://doi.org/10.1109/TSTE.2020.2966906.

[14] Shannon CE. A mathematical theory of communication. Bell Syst Tech J Jul. 1948; 27(3):379–423.

[15] Wang B, Guo Q, Yang T, Xu L, Sun H. Data valuation for decision-making with uncertainty in energy transactions: A case of the two-settlement market system. Appl Energy Apr. 2021;288:116643. https://doi.org/10.1016/j.apenergy.2021.116643.

[16] Stigler GJ. The Economics of Information. J Polit Econ Jun. 1961;69(3):213–65.

[17] Liang F, Yu W, An D, Yang Q, Fu X, Zhao W. A Survey on Big Data Market: Pricing, Trading and Protection. IEEE Access 2018;6:15132–54. https://doi.org/10.1109/ACCESS.2018.2806881.

[18] Exchange GBD. 2016 White Paper on China Big Data Exchange. China: Guiyang; May 2016.

[19] S. Liu, "Big data - Statistics & Facts." Aug. 2021. [Online]. Available: https://www.statista.com/topics/1464/big-data/.

[20] Oh H, Park S, Lee GM, Choi JK, Noh S. Competitive Data Trading Model With Privacy Valuation for Multiple Stakeholders in IoT Data Markets. IEEE Internet Things J. Apr. 2020;7(4):3623–39. https://doi.org/10.1109/JIOT.2020.2973662.

[21] Tian L, Li J, Li W, Ramesh B, Cai Z. Optimal Contract-Based Mechanisms for Online Data Trading Markets. IEEE Internet Things J. Oct. 2019;6(5):7800–10. https://doi.org/10.1109/JIOT.2019.2902528.
[22] Zhang Y, Niyato D, Wang P, Han Z. Data Services Sales Design With Mixed Bundling Strategy: A Multidimensional Adverse Selection Approach. IEEE Internet Things J. Sep. 2020;7(9):8826–36. https://doi.org/10.1109/JIOT.2020.2999824.
[23] Jiao Y, Wang P, Feng S, Niyato D. Profit Maximization Mechanism and Data Management for Data Analytics Services. IEEE Internet Things J. Jun. 2018;5(3): 2001–14. https://doi.org/10.1109/JIOT.2018.2819706.
[24] R. Cummings, K. Ligett, A. Roth, Z. S. Wu, and J. Ziani, "Accuracy for Sale: Aggregating Data with a Variance Constraint," in *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, New York, NY, USA, Jan. 2015, pp. 317–324. doi: 10.1145/2688073.2688106.
[25] A. Ghosh and A. Roth, "Selling privacy at auction," in *Proceedings of the 12th ACM conference on Electronic commerce*, New York, NY, USA, Jun. 2011, pp. 199–208. doi: 10.1145/1993574.1993605.
[26] A. Agarwal, M. Dahleh, and T. Sarkar, "A Marketplace for Data: An Algorithmic Solution," in *Proceedings of the 2019 ACM Conference on Economics and Computation*, New York, NY, USA, Jun. 2019, pp. 701–726. doi: 10.1145/3328526.3329589.
[27] Gonçalves C, Pinson P, Bessa RJ. Towards Data Markets in Renewable Energy Forecasting. IEEE Trans Sustain Energy Jan. 2021;12(1):533–42. https://doi.org/10.1109/TSTE.2020.3009615.
[28] Chen L, Wu Z, Wang J, Yu M, Yu Y, Li G, et al. *IEEE Power Energy Society General Meeting (PESGM)*, Washington D.C. USA, Jul. 2021;2021:1–5. https://doi.org/10.1109/PESGM46819.2021.9638205.
[29] B. Wang, Q. Guo, T. Yang, and H. Sun, "Cooperative Game Theoretic Framework for Electricity Retailers Participating in Data Transactions," in *2021 IEEE Power Energy Society General Meeting (PESGM)*, Washington D.C., USA, Jul. 2021, pp. 1–5.
[30] L. Han, J. Kazempour, and P. Pinson, "Monetizing Customer Load Data for an Energy Retailer: A Cooperative Game Approach," in *2021 IEEE Madrid PowerTech*, Jun. 2021, pp. 1–6. doi: 10.1109/PowerTech46648.2021.9494834.
[31] Rashidizadeh-Kermani H, Vahedipour-Dahraie M, Shafie-Khah M, Siano P. A Regret-Based Stochastic Bi-Level Framework for Scheduling of DR Aggregator Under Uncertainties. IEEE Trans Smart Grid Jul. 2020;11(4):3171–84. https://doi.org/10.1109/TSG.2020.2968963.
[32] Pinson P, Chevallier C, Kariniotakis GN. Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power. IEEE Trans Power Syst Aug. 2007;22 (3):1148–56. https://doi.org/10.1109/TPWRS.2007.901117.
[33] Matevosyan J, Soder L. Minimization of imbalance cost trading wind power on the short-term power market. IEEE Trans Power Syst Aug. 2006;21(3):1396–404. https://doi.org/10.1109/TPWRS.2006.879276.
[34] Tsimopoulos EG, Georgiadis MC. Optimal strategic offerings for a conventional producer in jointly cleared energy and balancing markets under high penetration of wind power production. Appl Energy Jun. 2019;244:16–35. https://doi.org/10.1016/j.apenergy.2019.03.161.
[35] Porteus EL. Foundations of Stochastic Inventory Theory. Stanford, CA, USA: Stanford University Press; 2002.
[36] Du E, Zhang N, Kang C, Kroposki B, Huang H, Miao M, et al. Managing Wind Power Uncertainty Through Strategic Reserve Purchasing. IEEE Trans Power Syst Jul. 2017;32(4):2547–59. https://doi.org/10.1109/TPWRS.2016.2617466.
[37] Benndorf V, Normann H. The willingness to sell personal data. Scand J Econ Oct. 2018;120(4):1260–78. https://doi.org/10.1111/sjoe.12247.
[38] Oh H, Park S, Lee GM, Heo H, Choi JK. Personal Data Trading Scheme for Data Brokers in IoT Data Marketplaces. IEEE Access 2019;7:40120–32. https://doi.org/10.1109/ACCESS.2019.2904248.
[39] Baeyens E, Bitar EY, Khargonekar PP, Poolla K. Coalitional Aggregation of Wind Power. IEEE Trans Power Syst Nov. 2013;28(4):3774–84. https://doi.org/10.1109/TPWRS.2013.2262502.
[40] Mas-Colell A, Whinston MD, Green JR. Microeconomic Theory. New York, NY, USA: Oxford University Press; 1995.