

Forecast Competition in Energy Imbalance Market

Jingshi Cui, *Graduate Student Member, IEEE*, Nan Gu, *Graduate Student Member, IEEE*, Tianyu Zhao, Chenye Wu, *Member, IEEE*, and Minghua Chen, *Senior Member, IEEE*

Abstract—Uncertainties in renewable generation make accurate load forecast essential for reliable power system operation. This paper considers energy imbalance market (EIM), where market players are allowed to procure energy ahead of time and trade the mismatch due to forecast error and strategic behaviors. The ISO sets the trading prices according to the market conditions, and pursues various system-level objectives. We first identify the power-law relationship between data volume and forecast accuracy, which enables the formulation of forecast cost model. Then, we cast the interactions in the EIM in the Stackelberg game framework with the ISO acting as the leader. We offer explicit subgame perfect equilibrium among the players in EIM, and derive the sufficient condition for the existence of unique equilibrium. Then, we show that this equilibrium, if exists, supports the maximal social welfare and under certain conditions, minimizes the total mismatch. We further examine the local and global impacts of the forecast errors under mild conditions, together with robustness analysis. Such analysis provides mechanism design guidelines for the ISO to enable the data sharing and forecast method sharing among market players in the EIM. Numerical studies further examine the effectiveness, robustness and sensitivity of the subgame.

Index Terms—Load Forecast, Stackelberg Game, Subgame Perfect Equilibrium, Power Law

NOMENCLATURE

Functions

$\tilde{B}_{(\cdot)}$	Benefit function with penalty term of each player.
$B(\cdot)$	Social welfare of the system.
$B_{(\cdot)}$	Benefit function of each player.
$C_d(\cdot)$	Generic forecast cost function.
$f_{e_i}(\cdot)$	PDF of player i 's forecast error.
$F_e(\cdot)$	CDF of total forecast error.

Indexes and Numbers

i	Index for player.
N	Number of players in EIM.

Parameters

λ_d	Unit price for each unit volume of data.
λ_e	Unit energy excess penalty.
λ_r	Energy trading price in the EIM.
λ_s	Unit energy shortage penalty.
λ_t	Energy procurement price ahead of time.
λ_u	Energy price for local users.
μ_i	Mean value of player i 's forecast error.
σ_i	Standard deviation of player i 's forecast error.
a, b, c	Parameters of relationship between data volume and forecast accuracy.
C_0	Fixed access fee of forecast training data.
p	Unit penalty coefficient in benefit function with penalty term.
P_i	Parameter of player i 's benefit function with penalty term to ensure budget balance.

Variables

α	Aggregate strategic behavior of all players.
α_i	Player i 's strategic behavior.
Δ	Total mismatch in the system.
Δ_i	Mismatch of player i due to forecast error and strategic behavior.
\hat{D}_i	Forecast net load of player i .
\hat{r}_i	Forecast renewable energy of player i .
D_i	True load of player i .
e	Total forecast error of all players.
e_i	Forecast error of player i .
L_i	Demand under the purview of player i .
Q_i	Energy procurement ahead of time for player i .
r_i	True renewable energy of player i .
v	Data volume of forecast model's training set.

I. INTRODUCTION

Classically, quantitative trading is one of the major means for institutional investment, which enabled the last round of trading revolution in the financial market. It utilizes advanced mathematical models and computer technology, to replace human subjective judgment. Quantitative trading reduces the impact of traders' sentiment by avoiding irrational decision making in extremely enthusiastic or pessimistic conditions. Advanced machine learning and artificial intelligence techniques are reshaping the landscape of financial market trading by introducing algorithmic trading. It shifts the focus of classical quantitative trading to price trend forecast and real-time decision making.

The development of electricity market introduces many forms of financial instruments for various purposes, e.g., competition promotion, or fluidity enhancement in the market. In the recent years, there is a similar trend of advancement

Manuscript received May 12, 2021; revised August 20, 2021; accepted September 26, 2021. Date of publication XXX, 2021; date of current version XXX, 2021. This work was supported in part by a Start-up Grant from School of Data Science (Project No. 9380118), City University of Hong Kong, a General Research Fund from Research Grants Council, Hong Kong (Project No. 11206821), and Shenzhen Institute of Artificial Intelligence and Robotics for Society. Paper no: TPWRS-00748-2021. *Corresponding author: Chenye Wu.*

J. Cui and N. Gu are with the Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084 China.

T. Zhao is with Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong.

C. Wu is with the School of Science and Engineering, the Chinese University of Hong Kong, Shenzhen, Shenzhen, Guangdong 518172 China, and also with the Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen, Guangdong 518129 China (email: chenye_wu@yeah.net).

M. Chen is with School of Data Science, City University of Hong Kong, Hong Kong.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier XXX

in the electricity market and the general financial market. Advanced machine-learning-based forecast methods enable the innovation of financial instruments in the electricity market.

A. Opportunities and Challenges

To date, major players in virtual bidding and financial transmission right markets have already developed their own machine-learning-based decision-making systems. In this paper, we focus on the interaction of heterogeneous decision-making systems in the energy imbalance market (EIM), firstly proposed by the California Independent System Operator (CAISO) in 2013 for coordinating the day-ahead market and the real-time market [1].

However, the interaction is complex. The first difficulty comes from the unknown cost structure of the forecast methods, e.g., what is the exact relationship between the data volume (or the running time) and the forecast accuracy? Even if the relationship is clear, it is imprecise how the forecast accuracy affects the players' profits in the EIM. This leads to the second difficulty, i.e., how to characterize the potential strategic behaviors of market players with advanced forecast methods? In other words, under which circumstances, the market reaches an equilibrium? These are the major focuses of our work. With the equilibrium characterization, we further analyze the market effectiveness at the equilibrium. Specifically, we identify the guidelines for the Independent System Operator (ISO) to enable information sharing (including both data and forecast method) between competitors.

B. Related Works

In tackling the aforementioned challenges, we identify two major streams of closely related works which investigate 1) the optimal design of various forecast methods in the electricity sector and 2) the impacts of forecast results on the electricity market outcomes.

There is a rapidly growing body of research on the optimal design of forecast methods in the electricity sector, from load forecast to electricity price forecast. For load forecast, Ghadimi *et al.* design a hybrid load forecast method containing smart feature selection based on Elman neural networks [2]. Bedi *et al.* propose a deep-learning-based framework to forecast electricity demand by combining cluster analysis, load trend characterization, and Long Short Term Memory (LSTM) network in [3]. Liu *et al.* design a hybrid forecast engine for small scale load forecast based on sliding window empirical mode decomposition in [4]. Many researchers extensively employ advanced learning frameworks for the electricity price forecast. For instance, Qiao *et al.* generate a hybrid model consisting of wavelet transform, stacked auto-encoder, and LSTM in [5]. Chitsaz *et al.* propose an intra-hour rolling horizon framework using high-resolution data for price forecast in [6]. Bello *et al.* forecast electricity prices in the medium term focusing on accurate estimates of tail risks in [7]. Gao *et al.* find out the correlation between load and price forecast, which leads to an improved forecast engine in [8]. The focus of latest researches goes beyond the value forecast of load or price. The focus shifts to (i) forecast probability distribution

function of the generated power in photovoltaic systems based on the higher-order Markov chain [9], (ii) a new prototype tool developed in Electric Reliability Council to forecast system inertia and adequacy evaluation of frequency response reserves [10], etc. In contrast to the above literature, whose goal is to improve the forecast accuracy, we intend to study how the data volume (and hence the cost of accessed data) may affect the forecast accuracy based on the existing forecast methods. Specifically, we seek to establish the functional relationship between data volume and forecast accuracy, which helps us uncover the true cost for the market players to obtain accurate forecast.

Uncertainty exists in many aspects of the electricity sector, such as uncertainty of cooling demand [11], peak load management of industrial consumers [12], etc. A well-investigated topic is to quantify uncertainties impacts (mostly due to forecast errors) on the electricity market outcomes. For example, Abedinia *et al.* consider uncertainties in demand, power prices and weather information (solar, wind, temperature, etc.) to obtain consumer's strategies in [13]. Sun *et al.* propose a probabilistic day-ahead net load forecast method to capture both epistemic uncertainty and aleatoric uncertainty using Bayesian deep learning in [14]. To quantify the uncertainty of future electricity demand, Taieb *et al.* estimate an additive quantile regression model for future distribution using a boosting procedure in [15]. Baker *et al.* provide a tool to optimally size energy storage systems to handle the difficulties brought by the renewable generation forecast errors in the model of predictive control framework in [16]. Xie *et al.* propose an energy storage capacity optimization method for grid-connected microgrid systems considering multiple time scale uncertainty coupling in [17]. In this paper, we focus on the impact of uncertainties in a specific market, the EIM: We characterize how the forecast competition between players affects the market outcome.

C. Our Contributions

Two closely related works inspire our work. The first work is by Zhao *et al.*, which studies the strategic bidding behaviors of players under a game-theoretical setting in the deregulated electricity market, taking the uncertainty in both demand and local renewable generation into account [18]. They assume that forecast error follows independent zero-mean Gaussian distribution, and model the interactions among players as a competitive game. Our new idea is to examine the local impact and global impact of forecast errors on the market outcome. Another work is by Kalathil *et al.*, which characterizes equilibrium prices for shared storage in a spot market, and formulates storage investment decisions of the firms as a non-convex competitive game in [19]. Our inspiration source for conducting this study is the solution concept in analyzing the sharing market; nevertheless, our work is fundamentally different from these two works in terms of both research scope and technical contributions.

To the best of our knowledge, we are the first to discuss forecast competition in the electricity sector. We summarize the principal contributions of our work as follows:

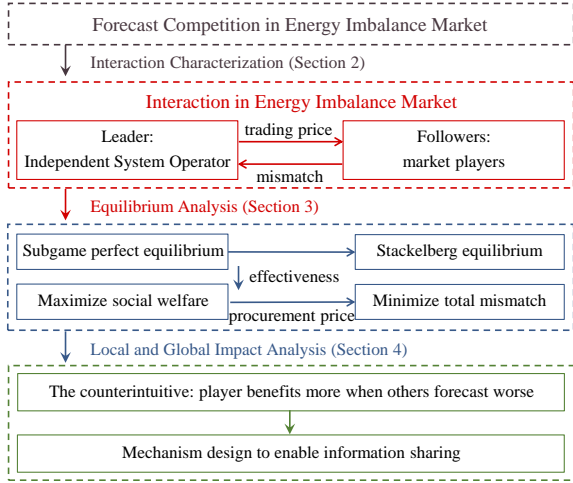


Fig. 1. Paradigm of Forecast Competition in EIM

- **Identification of Forecast Cost:** We first use numerical studies to highlight the power-law relationship between data volume and forecast accuracy (in terms of Mean Square Error (MSE)), and then establish the cost function for market players to obtain forecast results with guaranteed accuracy.
- **Equilibrium Analysis for Forecast Competition:** We study the Stackelberg equilibrium for the forecast competition and offer an explicit characterization of general subgame perfect equilibrium. We then propose the sufficient condition for its existence and uniqueness, and examine its robustness.
- **Local and Global Impact Analysis:** Assuming the forecast error distribution as independent zero-mean Gaussian, we investigate the local and global impact of forecast errors on the market outcome. Based on such analysis, we design a guideline for the ISO to achieve various system-level objectives. We also propose a mechanism for the ISO to enable information sharing among players.

The structure of this paper is as follows. We first introduce the forecast competition model in the EIM and construct a forecast cost model in terms of data volume in Section II. Based on the competition model, Section III formulates the EIM interaction game with subgame analysis and studies the effectiveness of the equilibrium. Section IV investigates the local and global impacts of forecast error on market outcome. In Section V, extensive numerical studies highlight the effectiveness of our proposed market. Finally, Section VI delivers the concluding remarks. Fig. 1 indicates the structural relationship between sections.

II. FORECAST COMPETITION MODEL

This section introduces the interaction between market players in the EIM, and then formally characterizes the benefit function for each player.

A. Interaction in EIM

We consider N players in the EIM, where each player could be a local utility company, or an aggregator. The player is

responsible for the stability of its region via participating in the EIM. Specifically, at each time period, player i needs to forecast its net load \widehat{D}_i :

$$\widehat{D}_i = L_i - \widehat{r}_i, \quad (1)$$

where L_i is the demand under the purview of player i , and \widehat{r}_i is the forecast renewable energy output. This paper takes wind energy as an example. Correspondingly, Eq. (2) defines the true (or realized) net load D_i :

$$D_i = L_i - r_i. \quad (2)$$

Obviously, there is a mismatch between D_i and \widehat{D}_i due to the renewable energy forecast error, denoted by the random variable of e_i :

$$e_i = r_i - \widehat{r}_i. \quad (3)$$

Its mean and variance are as follows:

$$\mathbb{E}\{e_i\} = \mu_i, \quad D\{e_i\} = \sigma_i^2. \quad (4)$$

Due to the forecast error, each player first procures certain energy ahead of time at price λ_i , and then purchases the deficit or sells the excess via the EIM. Note that in the first stage, the players may strategically over-purchase or under-purchase based on their own forecast knowledge. Mathematically, we denote the energy procurement ahead of time for player i by Q_i , which is different from \widehat{D}_i . Let α_i denote such difference by

$$\alpha_i = \widehat{D}_i - Q_i, \quad (5)$$

which measures player i 's strategic behavior.

At the system-level, in the EIM, the mismatch Δ_i is the result of player i 's forecast error and its strategic behavior. Specifically, the calculation of Δ_i is as follows,

$$\Delta_i = D_i - Q_i = \alpha_i - e_i. \quad (6)$$

The ISO maintains the EIM according to the total mismatch of Δ in the system, where $\Delta = \sum_{j=1}^N \Delta_j$. Specifically, it determines the trading price λ_r in the EIM as follows:

$$\lambda_r = \begin{cases} \lambda_s & \Delta > 0 \\ 0 & \Delta = 0 \\ \lambda_e & \Delta < 0 \end{cases} \quad (7)$$

where λ_s and λ_e are the unit energy shortage and excess penalty, respectively. In this paper, we assume that the utilities are price takers, and the shortage and excess penalty mainly reflect the cost of emergency power generation and disposal of excess electricity. Hence, we take λ_s and λ_e as fixed values; nonetheless, λ_s and λ_e can be different in various time slots¹. Note that we require $\lambda_e \leq \lambda_s$ because it is in general more expensive to handle the emergency energy shortage (i.e., to call for spinning and non-spinning reserve) in real-time than to handle the energy excess (i.e., to compensate the lost opportunity cost of certain generators).

Intuitively, $\Delta > 0$ means that the EIM requires additional emergency energy purchases. In this case, if $\Delta_i > 0$ for player i , energy purchase price is λ_s . Otherwise, player i sells its

¹While λ_s and λ_e can be time varying, we assume that they are decided ahead of time for single shot EIM interaction, which is the focus of our paper.

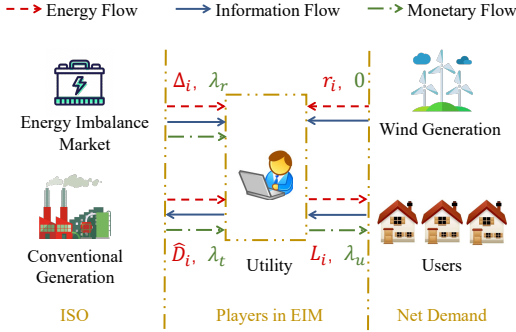


Fig. 2. Framework of EIM with Multiple players

excess energy at price λ_s , thereby earning through the price difference between λ_s and energy procurement price λ_t . When $\Delta < 0$, the analysis follows the same routine.

B. Model for Each Player

Fig. 2 shows the framework of EIM with multiple market players. The red, blue, and green arrows indicate the direction of energy, information, and monetary flow, respectively. The red variables are the quantity of energy; and the green parameters are the corresponding prices.

As indicated in Fig. 2, the benefit function of each player consists of four parts: energy procurement cost ahead of time at price λ_t , energy trading benefit from EIM at price λ_r , energy fee collection from local users at price λ_u ², and the cost to generate a reliable forecast. Mathematically, we denote the benefit function of player i by B_i :

$$B_i = -\lambda_t \cdot Q_i + \lambda_r \cdot (-\Delta_i) + \lambda_u \cdot L_i - C_d(MSE_i), \quad (8)$$

where λ_r is a random variable defined in Eq. (7); and MSE_i is player i 's forecast requirement (we select Mean Square Error, MSE in short, as the metric for forecast accuracy). For all the prices, it holds that:

$$0 < \lambda_e < \lambda_t \leq \lambda_u < \lambda_s. \quad (9)$$

This requirement is quite intuitive. To ensure the profit of the player, $\lambda_t \leq \lambda_u$. The emergency energy purchase should be at a higher cost as $\lambda_t \leq \lambda_u < \lambda_s$. Meanwhile, the lost opportunity cost to handle the energy excess should be less than the net cost of the generations, yielding $0 < \lambda_e < \lambda_t$.

The only remaining hurdle in modeling the benefit function for each player is to identify the form of forecast cost C_d in terms of forecast accuracy MSE_i , which is our next target.

C. Forecast Cost Formulation

In this digital age, data is an important asset. For example, Agarwal *et al.* introduce the auction mechanism for a two-sided data market in [20], where the data marketplace enables effective training data trading for machine learning tasks to achieve a certain level of accuracy. There is also a classical

²We mainly study the interaction among players in EIM. Hence, to simplify our analysis, we assume that users face fixed electricity retail rate.

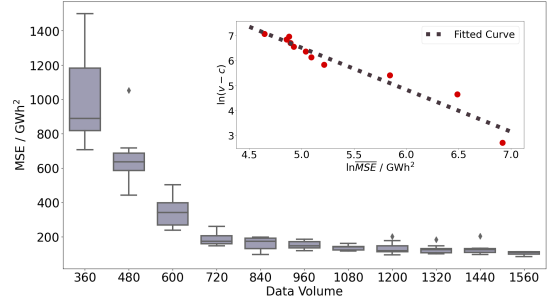


Fig. 3. Boxplot and Log-Log Graph: MSE v.s. Data Volume

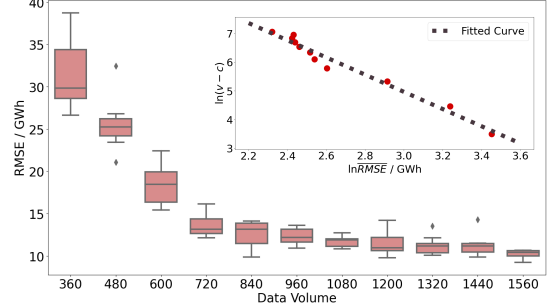


Fig. 4. Boxplot and Log-Log Graph: $RMSE$ v.s. Data Volume

way to study data cost and the sample complexity. Kearns first proposed this notion in the 1990s to study the minimum number of samples required to achieve learning objectives in the distribution-free model in [21]. From an information-theoretic approach, this line of research seeks to design various lower bounds for the “minimum number of required samples”. Given these two lines of research, we aim to study the relationship between data volume and forecast accuracy, which builds up the forecast cost formulation.

Remark 1: This approach is also very intuitive by considering the general underfitting and overfitting issues in machine learning. In case of limited data, the training model are incapable of fully capturing the key features in the data, which results in underfitting. In case of too much data, exceeding the model capacity, the resulting model may suffer from generalization inability, referred to as overfitting. Hence, the relationship between data volume and forecast accuracy is clear: the accuracy first increases with the growing data volume; then the increasing rate slows down, and it may even finally drop.

We use an empirical study to fit such a relationship. We select LSTM as the predictor and two metrics for forecast accuracy. Fig. 3 fits the relationship between training data volume and MSE , and Fig. 4 replaces metric MSE with Root Mean Squared Error ($RMSE$). The fitted curves are consistent with many of our intuitions: if error is already small, the marginal value of data is very limited; and if the available data volume is limited, it is very beneficial to purchase more data. In fact, we find that the data volume and forecast accuracy follow a power-law relationship. From an intuitive point, when the forecast accuracy is already high, it is difficult to increase the accuracy rate. Conversely, when the forecast accuracy is very low, adding a small amount of data

can increase the accuracy dramatically. We also replot log-log graph with average value in Fig. 3 and Fig. 4. The R^2 values of 0.94 (MSE) and 0.97 ($RMSE$) verify the strong power-law relationship.

Remark 2: In this paper, our forecast model is LSTM, and the metric for forecast accuracy is MSE . However, we emphasize that the power-law relationship between data volume and forecast accuracy is quite general as we observed similar results using data from NordPool and PJM.

Thus, the relationship between data volume v and MSE is as follows:

$$v = a \cdot MSE^{-b} + c, \quad (10)$$

where parameters a and b are both positive.

Since the data do not come free, we assume a generic cost function $C_d(v)$ to gain the access to a dataset of volume v :

$$C_d(v) = C_0 + \lambda_d \cdot v, \quad (11)$$

where C_0 is the fixed access fee, and λ_d is the unit price for each unit volume of data.

Together with Eq. (10), the forecast cost function in terms of forecast accuracy MSE is:

$$C_d(MSE) = C_0 + \lambda_d \cdot c + \lambda_d \cdot a \cdot MSE^{-b}. \quad (12)$$

III. EQUILIBRIUM ANALYSIS FOR EIM

Based on the forecast competition model, we formulate a Stackelberg game to fully capture the interaction between the ISO (market organizer) and all the market players in the EIM:

EIM Interaction Game (EIG):

Players: ISO acts as leader and market players in EIM act as followers;

Strategies: ISO sets the energy trading price λ_r (specifically λ_s and λ_e); market players select their strategic energy procurement α_i 's according to their forecast information.

Benefit Functions: The ISO may have a range of benefit functions from maximizing the social welfare to minimizing the total mismatch. We discuss how these diverse benefit functions may shape the equilibrium. Eq. (8) is the benefit function B_i for each market player.

We refer to this Stackelberg game as the EIM Interaction Game (EIG). The general solution for Stackelberg equilibrium is backward induction: we characterize the subgame perfect equilibrium between all the market players given λ_r , and then utilize the subgame perfect equilibrium to optimally determine λ_r for establishing the equilibrium of EIG.

Before diving into detailed equilibrium analysis, we first make the following assumption on forecast error distribution:

Assumption 1: Denote the probability density function of forecast error e_i by $f_{e_i}(x)$. We assume $f_{e_i}(x) > 0, \forall x \in \mathbb{R}, \forall i$.

Remark 3: This is a standard assumption in analyzing the forecast error. The forecast error is often assumed to follow various continuous distributions, such as Gaussian distribution [22], Lévy α -stable distribution [23], etc, where this

assumption automatically holds. In practice, we often don't need to strictly require $f_{e_i}(x)$ is positive over \mathbb{R} . Instead, it often suffices for $f_{e_i}(x)$ to be positive over some continuous interval.

A. Subgame Analysis

This assumption paves the way for further characterizing the subgame perfect equilibrium. The key is to examine the first-order optimality condition for all the benefit functions B_i 's:

$$\frac{\partial \mathbb{E}\{B_i\}}{\partial \alpha_i} = 0. \quad (13)$$

All the forecast errors e_i 's are the expected values as other players' forecast errors also affect player i 's benefit function through λ_r . This also highlights the coupling between players in the EIG.

Theorem 1: If the subgame of EIG admits an equilibrium, denoted by $\{\alpha_i^*, \forall i\}$, the equilibrium is the unique subgame perfect equilibrium for given λ_e and λ_s , characterized as follows:

$$\alpha_i^* = \mathbb{E}\{e_i | e = \alpha^*\}, \quad (14)$$

$$\alpha^* = F_e^{-1}\left(\frac{\lambda_t - \lambda_e}{\lambda_s - \lambda_e}\right), \quad (15)$$

where $e = \sum_{j=1}^N e_j$; $\alpha^* = \sum_{j=1}^N \alpha_j^*$; and $F_e(\cdot)$ is the cumulative distribution function of e .

Remark 4: This result follows the classical result of the newsvendor problem [24]. The detailed proof of Theorem 1 is given in Appendix A. However, Theorem 1 does not guarantee the existence of subgame perfect equilibrium. This is due to the non-convexity of the game formulation. We exemplify the fact that equilibrium may not always exist as follows.

Example 1: Consider two market players in the EIM. Their respective forecast error distributions are as follows:

$$e_1 = \rho^2, \quad e_2 = \rho - \rho^2, \quad (16)$$

where ρ follows uniform distribution $U(-10, 10)$. Set $\lambda_e = 1$, $\lambda_t = 2$, and $\lambda_s = 3$. According to Theorem 1, if the equilibrium exists, then it must be achieved at $\alpha_1^* = 0$ and $\alpha_2^* = 0$.

However, we can further examine the first-order optimality condition for the first player:

$$\begin{aligned} \frac{\partial \mathbb{E}\{B_1\}}{\partial \alpha_1} &= \lambda_t - \lambda_e - (\lambda_s - \lambda_e)F_e(\alpha) \\ &\quad - (\lambda_s - \lambda_e)f_e(\alpha)(\alpha_1 - \mathbb{E}\{e_1 | e = \alpha\}) \\ &= \frac{1}{10}(\alpha_1^2 - 2\alpha_1), \end{aligned} \quad (17)$$

where the hypothetical equilibrium point ($\alpha_1^* = 0, \alpha_2^* = 0$) is a local maximum of B_1 . Hence, it is not the optimal choice for player 1. We use Fig. 5 to visualize the observations. The red point corresponding to $\alpha_1 = 0$ is the local maximum, and the blue point corresponding to $\alpha_1 = 2$ is the local minimum. Fig. 5(b) confirms that the local maximum corresponding to $\alpha_1 = 0$ is not the global maximum.

Thus, the existence of subgame perfect equilibrium needs the sufficient condition:

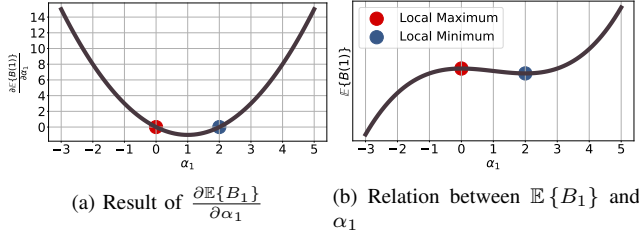


Fig. 5. Schematic Diagram of Example

Theorem 2: If the following technical alignment condition holds,

$$\frac{\partial \mathbb{E}\{e_j | e = \rho\}}{\partial \rho} \geq 0, \quad \forall j, \quad (18)$$

the subgame perfect equilibrium characterized in Theorem 1 uniquely exists.

The key to prove this theorem is to use condition (18) to examine the monotonicity of the first order condition for each benefit function. Appendix B provides detailed proof.

Remark 5: It is easy to satisfy this technical alignment condition. It merely requires that when the forecast becomes hard (e.g., due to an unexpected holiday, a storm in the summer, etc.), it is hard for every player. Example 1 fails to admit an equilibrium exactly due to the violation of the technical alignment condition:

$$\frac{\partial \mathbb{E}\{e_2 | e = \rho\}}{\partial \rho} = 1 - 2\rho, \quad (19)$$

which might be negative over $[-10, 10]$.

B. Equilibrium Effectiveness

The ISO seeks to achieve maximal social welfare in terms of maximizing the total benefit functions of all players. The ISO also sets the trading prices in EIM (λ_s and λ_e) to minimize the total mismatch in the EIM, for balancing the supply and demand in real-time.

Denote the social welfare of the system by B :

$$B = \sum_{j=1}^N B_j. \quad (20)$$

We can show that, no matter how the trading prices are set in EIM, the subgame perfect equilibrium automatically supports the maximal social welfare:

Theorem 3: If the subgame perfect equilibrium $\{\alpha_i^*, \forall i\}$ holds, it supports the maximal social welfare, i.e.,

$$\alpha^* = \arg \max_{\alpha} \mathbb{E}\{B\}. \quad (21)$$

It can be proved by considering the case where all players are merged into a single player, and showing the subgame perfect equilibrium coincides with the optimal decision making for the single player, as dictated in Theorem 1. The proof is demonstrated in Appendix C.

Remark 6: The maximal social welfare is a function of the trading prices (λ_s and λ_e). Hence, the ISO still could select the proper prices to influence the exact value of social welfare.

This paper also examines the ISO's strategy if it minimizes the total mismatch, in addition to maximization of social welfare. Specifically, it adopts the absolute value of mismatch in the entire EIM, $|\Delta|$, as the metric.

Theorem 4: Under Assumption 1 and technical alignment condition, if the ISO sets the prices λ_e and λ_s such that

$$\lambda_e + \lambda_s = 2\lambda_t, \quad (22)$$

the ISO could also minimize the total mismatch in EIM:

$$\alpha^* = \arg \min_{\alpha} \mathbb{E}\{|\Delta|\}. \quad (23)$$

This theorem can be proved by checking the partial derivative of $\mathbb{E}\{|\Delta|\}$ with respect to α . The detailed proof is given in Appendix D.

Remark 7: It is difficult to satisfy condition (22). For example, during peak demand seasons, the cost to call for emergency supply can be far more expensive than the day-ahead price. This partially proves that designing a fully competitive electricity market is very challenging without any manipulation, due to the complex structure of the power system.

C. Equilibrium Robustness

At equilibrium, each player adopts its optimal strategy in response to all other players executing their specified strategies. In practice, when some players are irrational or adversarial, there could exist fault behaviors which are not optimal given others' strategies. It could be an issue if these players can take advantage of the fault or adversarial behaviors for more benefit. This is often referred to as the robustness analysis. For our game formulation, we can obtain Theorem 5.

Theorem 5: Under Assumption 1 and technical alignment condition, the subgame perfect equilibrium $\{\alpha_i^*, \forall i\}$ is $(0, N-1)$ fault immune. That is, consider all players except a group of players S , $S \subset \{1, 2, \dots, N\}$ and $1 \leq |S| \leq N-1$. Given $\alpha_j = \alpha_j^*$, $\forall j \in \{1, 2, \dots, N\} \setminus S$, $\mathbb{E}\{B_j\}$ takes minimum when $\alpha_S = \alpha_S^*$, where $\alpha_S = \sum_{i \in S} \alpha_i$ and $\alpha_S^* = \sum_{i \in S} \alpha_i^*$. In addition, $\mathbb{E}\{B_j\}$ is non-decreasing (respectively non-increasing) w.r.t. α_S when $\alpha_S > \alpha_S^*$ (respectively $\alpha_S < \alpha_S^*$).

The proof follows similar routine as the robustness analysis in [25]. Appendix E provides the complete proof.

Remark 8: This theorem demonstrates the robustness of the subgame perfect equilibrium $\{\alpha_i^*, \forall i\}$ in response to non-optimal strategies of a group of players. The non-optimal fault behaviors of irrational players will not decrease the benefit of other rational players.

IV. FORECAST COMPETITION IN EIM

Up to now, this paper is out of focus on how the forecast accuracy may affect each market player's strategy since the equilibrium characterization only concerns the first-order moment (i.e., expectation) while *MSE* deals with the second-order moment (i.e., variance). This section studies a practical scenario: In addition to benefit maximization (after the subgame perfect equilibrium is reached), the market players may

also want to minimize the variance of the benefit function to firmly secure the benefit.

For the ease of analysis, we assume that all the forecast errors follow the independent zero-mean Gaussian distributions, i.e., $e_i \sim N(0, \sigma_i^2), \forall i$, with possibly distinct variances. In this setting, MSE coincides with the variance:

$$MSE_i = \sigma_i^2. \quad (24)$$

Specifically, we can characterize the expectation and variance of benefit function B_i 's in terms of MSE_i and MSE_{-i} ($MSE_{-i} = \sum_{j=1, j \neq i}^N MSE_j$) as follows:

Proposition 1: If $e_i \sim N(0, \sigma_i^2), \forall i$, and the forecast errors are independent, we have

$$\mathbb{E}\{B_i\} = \lambda_u L_i - \lambda_t D_i - C_0 - \lambda_d c - \lambda_d a MSE_i^{-b} - \frac{(\lambda_s - \lambda_e) MSE_i}{\sqrt{2\pi} \sqrt{MSE_i + MSE_{-i}}}, \quad (25)$$

$$D\{B_i\} = \frac{1}{2} MSE_i (\lambda_e - \lambda_t)^2 + (\lambda_s - \lambda_t)^2 - \frac{(\lambda_e - \lambda_s)^2 MSE_i^2}{2\pi (MSE_i + MSE_{-i})}. \quad (26)$$

The proof for this proposition involves utilizing the properties of Gaussian distribution and advanced calculus techniques. Appendix F illustrates detailed calculation. Such characterizations pave the way for studying the impact of MSE_i and MSE_{-i} on the benefit function, specifically, the local and global impact of forecast errors [26].

Theorem 6: When MSE_{-i} is fixed, $D\{B_i\}$ increases monotonically with respect to MSE_i .

The proof requires solving the partial derivative of $D\{B_i\}$ with respect to MSE_i , and theoretical proof is given in Appendix G.

Remark 9: This measures the incentive of each market player to improve its forecast accuracy, as an improved forecast method would reduce the variance of its benefit function. This is the local impact of forecast accuracy, which is quite straightforward.

Theorem 7: When MSE_i is fixed, $\mathbb{E}\{B_i\}$ and $D\{B_i\}$ both increase monotonically with respect to MSE_{-i} .

The proof of monotonicity of $\mathbb{E}\{B_i\}$ with respect to MSE_{-i} is straightforward. It immediately follows Eq. (25) in Proposition 1. To prove the monotonicity of $D\{B_i\}$ with respect to MSE_{-i} , we only need to examine the partial derivative of $D\{B_i\}$ with respect to MSE_{-i} . For more details of the proof, please refer to Appendix H.

Remark 10: This is a quite interesting yet counter-intuitive result. When the forecast method of player i is fixed, the player can benefit from other players' inability to predict well, though at the cost of the larger variance of benefit function. This is the global impact of forecast accuracy, as this result specifies how one player's forecast method may affect the others' benefits.

These two theorems (especially Theorem 7) motivate us to design a mechanism for the ISO to enable the data sharing or even the forecast method sharing between market players. To overcome the global impact in Theorem 7, we introduce a

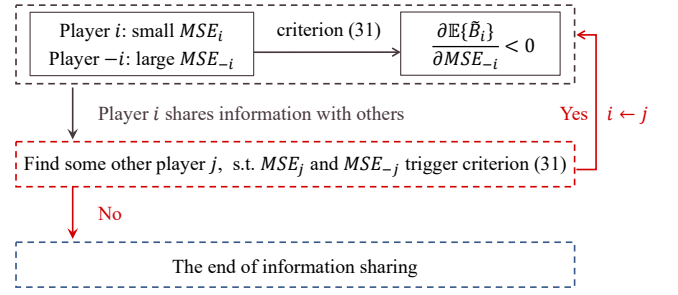


Fig. 6. Virtuous Cycle of Forecast Accuracy

penalty term to the benefit function B_i when the performances of forecast methods are too diverse in the system:

$$\begin{aligned} \tilde{B}_i = & -\lambda_t \cdot Q_i + \lambda_r \cdot (-\Delta_i) + \lambda_u \cdot L_i - C_d(MSE_i) \\ & + P_i - p \cdot \left(MSE_i - \frac{MSE_i + MSE_{-i}}{N} \right)^2, \end{aligned} \quad (27)$$

where \tilde{B}_i denotes the refined benefit function; p is the unit penalty coefficient; and P_i is a constant to ensure budget balance.

The penalty term affects the mean of the benefit functions:

$$\begin{aligned} \mathbb{E}\{\tilde{B}_i\} = & \mathbb{E}\{B_i\} + P_i + \frac{2p(N-1)}{N} MSE_i \cdot MSE_{-i} \\ & - \frac{p(N-1)^2}{N} (MSE_i^2 + MSE_{-i}^2). \end{aligned} \quad (28)$$

To highlight how the penalty term contains the global impact of MSE_{-i} , we need to examine the first-order derivative of $\mathbb{E}\{\tilde{B}_i\}$ with respect to MSE_{-i} :

$$\begin{aligned} \frac{\partial \mathbb{E}\{\tilde{B}_i\}}{\partial MSE_{-i}} = & \frac{(\lambda_s - \lambda_e) \cdot MSE_i}{2\sqrt{2\pi} (MSE_i + MSE_{-i})^{\frac{3}{2}}} \\ & + \frac{2p(N-1)}{N} MSE_i - \frac{2p}{N} MSE_{-i}. \end{aligned} \quad (29)$$

Standard mathematical manipulation yields that

$$\begin{aligned} & \frac{\partial \mathbb{E}\{\tilde{B}_i\}}{\partial MSE_{-i}} \\ & < \frac{\lambda_s - \lambda_e}{2\sqrt{2\pi} \left(1 + \frac{MSE_{-i}}{MSE_i}\right)} \\ & \quad + \frac{2p(N-1)}{N} MSE_i - \frac{2p}{N} MSE_{-i} \\ & < \frac{\lambda_s - \lambda_e}{2\sqrt{2\pi}} + \frac{2p(N-1)}{N} MSE_i - \frac{2p}{N} MSE_{-i}. \end{aligned} \quad (30)$$

The first inequality is due to $MSE_i + MSE_{-i} > 1$. The second is due to the positive sign of MSE_i and MSE_{-i} .

Hence, with the penalty term, we know that if

$$MSE_{-i} > \frac{N(\lambda_s - \lambda_e)}{4p\sqrt{2\pi}} + (N-1)MSE_i, \quad (31)$$

we have

$$\frac{\partial \mathbb{E}\{\tilde{B}_i\}}{\partial MSE_{-i}} < 0. \quad (32)$$

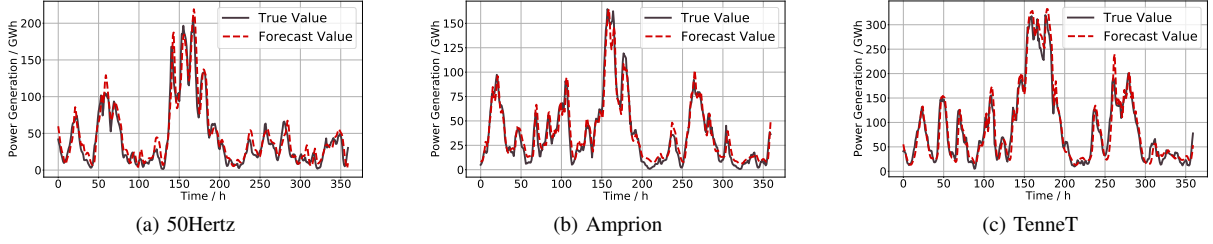


Fig. 7. Forecast Results of Three players

In other words, when player i 's own forecast results are good, but the average forecast results of other players in the system are too poor, the expected (refined) benefit of player i will decrease. This situation promotes the sharing of information (including data and forecast methods) among players in the EIM, so as to jointly improve the accuracy of forecast. Fig. 6 shows the underlying logic for the information sharing in the EIM.

V. NUMERICAL STUDIES

This section conducts extensive numerical studies to support our theoretical analysis, with special focus on the characterization of local and global impact.

A. Dataset Overview

We adopt two datasets: the wind power generation and the aggregate demand. Together, they allow us to estimate the market players' ability in forecasting the net demand. We consider three major players in Germany: 50Hertz, Amprion, and TenneT. For the convenience of representation, we use player 1, 2, and 3 to represent 50Hertz, Amprion, and TenneT, respectively. The wind power generations are provided in [27], collected from Jan. 1, 2016 to Apr. 14, 2016, with a resolution of 15 minutes. We aggregate the data to obtain the hourly wind power generation data, in alignment with the hourly energy consumption data. Such hourly consumption data for the three areas are in [28]. We select these two datasets as the wind power generation dataset contains diverse features beyond the wind power generation, including wind speed, temperature, and air pressure at the surface (as indicated in [29]). We can use such features to train the predictions and to obtain the cost parameters for achieving certain forecast accuracy.

We would also like to clarify the parameter selections in the numerical study. We adopt the energy retail price³ from [30] and set λ_u to be 72.9€/MWh. According to ISE (Fraunhofer Institute for Solar Energy Systems) 2013 [31], we set the energy procurement price λ_t to be 66.5€/MWh, the shortage penalty to be 80€/MWh, and the excess penalty to be 53€/MWh. As for the unit price for accessing data, we set the fixed cost C_0 to be 10,000€, and the price for each unit volume of data λ_d to be 70€.

The whole dataset is divided into two sets: the training set from Feb. 10, 2016 to Mar. 30, 2016, and the test set

³Note that in the Germany electricity market, various taxes and surcharges contribute roughly three quarters of the retail price. In our work, we only consider the pure retail electricity price without any tax or surcharge.

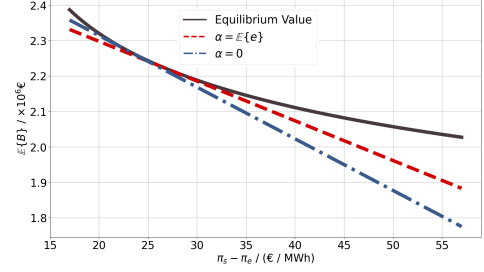


Fig. 8. Influence of Parameters and α on Social Welfare

from Mar. 31, 2016 to Apr. 14, 2016. We adopt the LSTM based predictor (see Appendix I for more details) and highlight the forecast performance for the three players in Fig. 7. The performance coincides with our expectations: the forecast value (represented by the red line) is most often consistent with the ground truth (represented by the black line). However, at breaking point (local minima and maxima), the forecast error is often large. This is the common issue for (net) load forecast [32].

B. Effectiveness, Sensitivity and Robustness Analysis

Section III investigates the effectiveness of equilibrium, and proves that the equilibrium α^* supports the maximal social welfare in Theorem 3. This section examines how the interaction game improves social welfare by comparing three strategies. The first one is the equilibrium behavior (illustrated by the black line in Fig. 8); the second one is to simply purchase the expected energy consumption without taking into account others' information (shown by the red line in Fig. 8); and the last strategy is to depart from the EIM, i.e., to set α_i to be 0 (indicated by the blue line in Fig. 8).

Fig. 8 compares the performance in terms of the social welfare achieved by the three strategies and by examining the trend of total social welfare with respect to the price difference between λ_s and λ_e (i.e., the arbitrage opportunities). Clearly, our equilibrium strategy achieves the best performance, which coincides with the conclusion in Theorem 3. Compared with the second strategy, we can observe the value of information. The information of all players improves the total social welfare by 6.8% when the arbitrage opportunity is high ($\lambda_s - \lambda_e = 55\text{€/MWh}$). Note that the second strategy does not always outperform the third strategy. When the arbitrage opportunity is high, even without the full information of the market, all the players may be willing to participate. However, when the arbi-

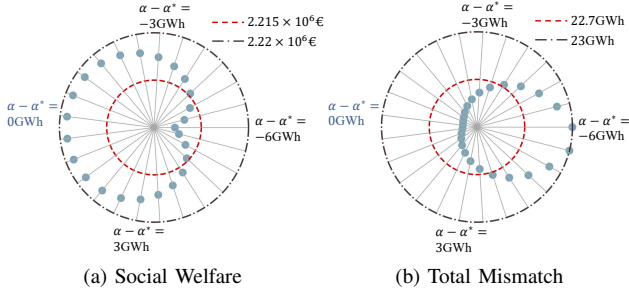


Fig. 9. Sensitivity Analysis on Social Welfare and Total Mismatch

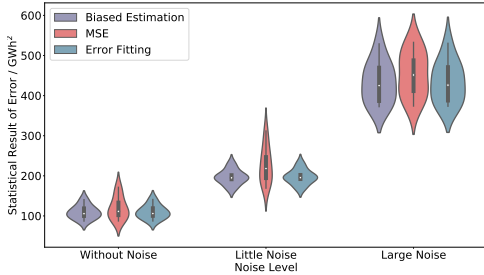


Fig. 10. Statistical Result of Forecast Error with Noise

trage opportunity is low (in our example, $\lambda_s - \lambda_e$ is less than 25€/MWh), many players may choose not to participate in the EIM. Such observations indicate the necessity of carefully designing the EIM.

To examine the sensitivity of subgame equilibrium, we consider two aspects. On the one hand, in Fig. 9, from the ISO's perspective, we study the impact of changes in α on expectation of social welfare $\mathbb{E}\{B\}$ and total mismatch $\mathbb{E}\{|\Delta|\}$. On the other hand, in Fig. 10, from the perspective of a single player, taking 50Hertz as an example, we explore the impact of noise in training set on the forecast error statistics.

Fig. 9(a) shows the changes of expected social welfare $\mathbb{E}\{B\}$ when $\alpha - \alpha^*$ changes from -6 GWh to 6 GWh. The red and black dotted lines represent $2.215 \times 10^6 \text{€}$ and $2.22 \times 10^6 \text{€}$, respectively. It is obvious from Fig. 9(a) that $\mathbb{E}\{B\}$ reaches its maximum when $\alpha = \alpha^*$, and as α deviates farther from α^* , $\mathbb{E}\{B\}$ decreases at a faster rate. This result is consistent with Theorem 3. Similarly, Fig. 9(b) describes the evolution of expected total mismatch, where the red and black dashed lines correspond to 22.7 GWh and 23 GWh, respectively. Consistent with Theorem 4, $\mathbb{E}\{|\Delta|\}$ reaches the minimum when $\alpha = \alpha^*$, and increases at a faster rate when α deviates farther from α^* .

In order to characterize the impact of noise on the sensitivity of the forecast model, we use three indexes to measure the forecast error: the biased variance estimation (empirical variance), MSE value, and the variance of the fitted Gaussian distribution. We select three cases for the training set: without noise, mixture with little noise (a noise following $U(-0.3, 0.3)$), and mixture with large noise (a noise following $U(-0.6, 0.6)$). We forecast multiple times and perform statistical analysis on forecast error. From Fig. 10, it is obvious that the increase in noise leads to an increase in forecast error, and

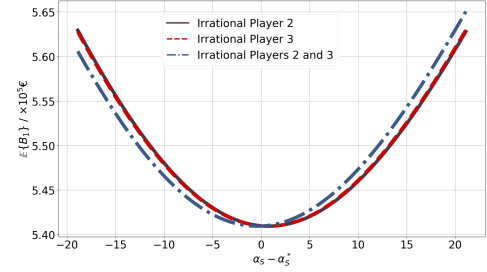


Fig. 11. Robustness of Subgame Perfect Equilibrium

a more discrete distribution of the statistical result of MSE , which reduces the feasibility and rationality of MSE .

To verify the robustness of subgame perfect equilibrium, we compare the performance of three cases in the general setting (i.e., error expectation is non-zero and players are dependent) in Fig. 11. The first two cases examine the performance when only one single player is irrational. The last case examines the performance when both player 2 and player 3 are irrational. In such cases, the increase of α_S of irrational player(s) deviating from the equilibrium value α_S^* causes the larger expected benefit of rational player 1, i.e., $\mathbb{E}\{B_1\}$. This result is guaranteed by Theorem 5. Taking the last case as an example, combined with the result of Fig. 9(a), we find that when $|\alpha_S - \alpha_S^*|$ is larger, the social welfare becomes smaller, and the benefit of the rational player increases. This illustrates that the more the irrational player deviates from the equilibrium, the more its benefit decreases. This observation promotes all players to follow the equilibrium action.

C. Gaussian Assumption Justification

We use the available data to justify our independent zero-mean Gaussian distribution assumptions for the forecast error. Fig. 12 plots the empirical forecast error distribution and its associated fittings. According to the literature, we select two fitting models: the Gaussian and the Levy α -stable distribution. Clearly in our work, the empirical forecast error distribution does not exhibit the heavy tail property; hence Gaussian distribution fits the empirical distribution better. In addition, all of the three players' empirical expectations are around zero, which further justifies our assumption is practical.

In Section IV, we submit that assuming Gaussian distributions, the biased variance estimation equals to MSE (see Eq. (24)). Similar to the three indexes in Fig. 10, Table I compares the biased variance estimation, MSE value, and the variance of the fitted Gaussian distribution. The three values are roughly the same, partially implying the validity of Gaussian distribution assumption.

The last justification is the assumption that the error distributions are independent. We study the correlation between error distributions by examining the correlation coefficient between the three players. Numerical calculation suggests that the weakest correlation is between Amprion and TenneT, equal to 0.1468; and the strongest one is between 50Hertz and TenneT, equal to 0.3051. Thus, the correlation is in general rather weak, which further justifies our assumption.

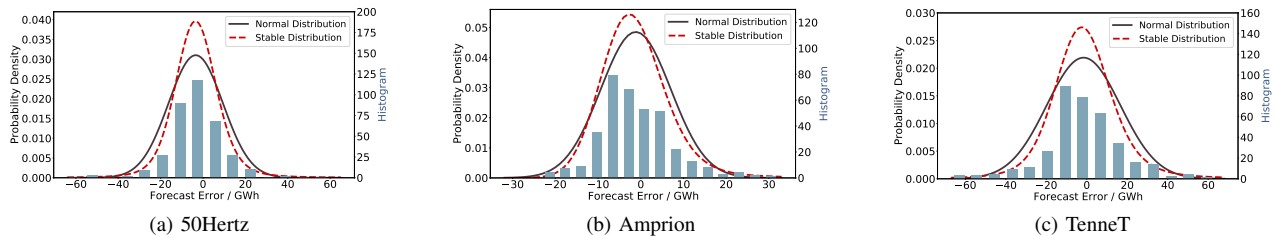
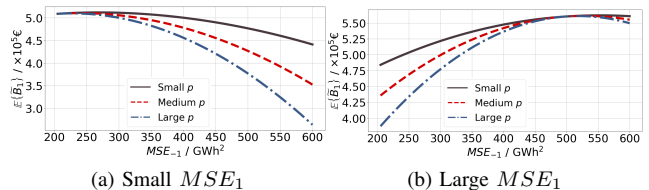
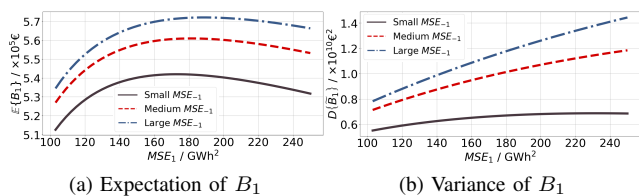


Fig. 12. Forecast Error Histogram and Probability Density Function Fitting

TABLE I
COMPARISON OF BIASED ESTIMATION VARIANCE, MSE AND ERROR FITTING VARIANCE (GWh^2)

Player	Biased Estimation	MSE	Error Fitting
50Hertz	165.0226	176.8760	165.4825
Amprion	67.4165	68.8016	67.6042
TenneT	330.6231	333.1033	331.5422

Fig. 14. Global Impact Considering Expectation of Improved Benefit Function with Small, Medium and Large p Fig. 13. Local Impact Considering Expectation and Variance of Benefit Function with Small, Medium and Large MSE_{-1}

D. Local and Global Impact

We use the data at 7 a.m. on Mar. 31, 2016 to examine the benefit of 50Hertz, highlighting the local and global impacts of forecast error. To align with our theoretical analysis, we refer to 50Hertz as player i , and the other two as players $-i$.

We first examine the local impacts by considering three cases for MSE_{-i} : small MSE_{-i} (200 GWh^2), true MSE_{-i} (401.9 GWh^2), and large MSE_{-i} (600 GWh^2). For all the three cases, Fig. 13(a) illustrates that the relationship between expected benefit and MSE_i share similar pattern. Fig. 13(b) illustrates the monotonic relationship between $D\{B_i\}$ and MSE_i . While reducing MSE_i may not always help improve the expected benefit, for fixed MSE_i , the larger MSE_{-i} is, the larger the expected benefit $\mathbb{E}\{B_i\}$ is. The observation reflects the global impacts in Theorem 7.

Also, we analyze the global impact with small penalty coefficient p ($p = 5$), medium p ($p = 10$), and large p ($p = 15$). Fig. 14(a) and Fig. 14(b) consider the impact of MSE_i on $\mathbb{E}\{B_i\}$ when MSE_i is small ($MSE_i = 100 \text{ GWh}^2$) and large ($MSE_i = 250 \text{ GWh}^2$), respectively. Both subfigures show that the penalty coefficient determines the increasing or decreasing rate of $\mathbb{E}\{B_i\}$. Comparing Fig. 14(a) and 14(b), an interesting result is that when MSE_{-i} is small, the increase of MSE_{-i} causes the decrease of $\mathbb{E}\{B_i\}$. On the other hand, when MSE_{-i} is large, the increase of MSE_{-i} causes increase of $\mathbb{E}\{B_i\}$. This result implies that all players in EIM tend to achieve the same level of forecast accuracy, which enables the

information sharing.

E. Dynamic Process of EIM

To study the decision making of all players in the system, Fig. 15 visualizes the payoff matrices for all the three players where grey, red and blue colors represent the three players of 50Hertz, Amprion, and TenneT, respectively. Here again, the time is at 7 a.m. on Mar. 31, 2016 for experimentation. The MSE_i of each player takes three values: small, medium, and large. The exact value is on the top of each subfigure, and the scattered points size of each color is under the same measurement, representing the size of $\mathbb{E}\{\tilde{B}_i\}$. For example, comparing the size of the scatter points in the three subgraphs in the first line, we observe the value of $\mathbb{E}\{\tilde{B}_1\}$ when MSE_1 , MSE_2 and MSE_3 take small, medium, and large values respectively (corresponding to $3 \times 3 \times 3 = 27$ combinations). Hence, the three subfigures together serve as payoff matrix in game theory.

Combining the results of 9 subfigures illustrates a dynamic adjustment process in the entire EIM. First, we assume that initially all the players do not spend too much effort on the forecast and they have large MSE s. In this case, player 3 first moves, as it observes that improving its forecast accuracy could significantly improve its benefit. Hence, player 3 reduces its MSE to 160 GWh^2 . The state transits from position 1 to position 2; and this state makes player 1 reconsider its situation. For player 1's decision making, we need to locate the same condition for player 1, moving from position 2 to position 3. Clearly, player 1 always would like to improve its accuracy. However, such a task could be hard for player 2 and it only reduces the MSE to the medium level, 175 GWh^2 . The state now transits from position 3 to position 4; and we can then consider player 2's decision making by further jumping to position 5. This procedure highlights competition in EIM.

However, information sharing could also achieve similar performance. For example, if we consider the initial benefit

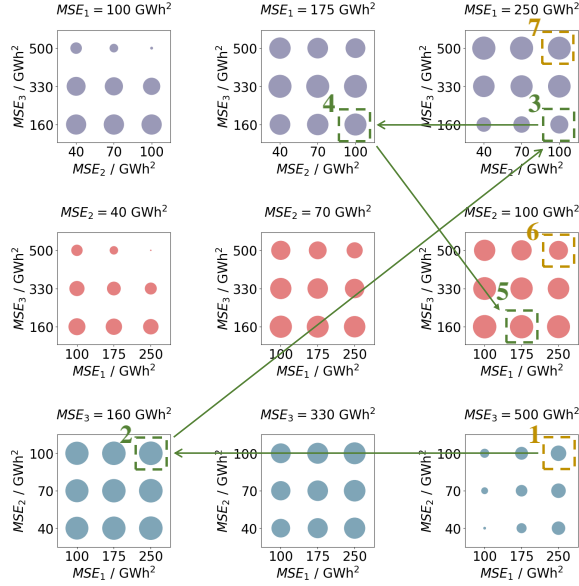


Fig. 15. Sample Dynamic Process of EIG

of the three players, highlighted by the yellow boxes (position 1, 6, and 7), it is clear that player 2 has both the incentive and the ability to help player 3 improve the forecast accuracy.

VI. CONCLUSION

In this work, we study the role of forecast accuracy in the benefit function of market players in the EIM. We further identify the power-law relationship between forecast accuracy and the cost of purchasing various volumes of data, which serves as the basis of player's benefit function formulation. This allows us to characterize the Stackelberg game interaction in the EIM. To examine the Stackelberg equilibrium behavior, we first offer the explicit characterization of the subgame perfect equilibrium with the sufficient condition for the unique existence of such equilibrium, which is supported by both complete theoretical proof and an intuitive counter example. We submit that the Stackelberg equilibrium supports maximal social welfare. To harness the benefit of all the players, in addition to the expected benefit, we also consider the variance of benefit. Under the independent zero-mean Gaussian distribution assumption for all forecast errors, we investigate the local and the global impacts of each individual player's forecast accuracy on the final outcome. We provide a guideline for ISO as an example by introducing a penalty term to enable the information sharing among the competitors in the EIM.

The numerical studies examine the effectiveness, sensitivity and robustness of the subgame equilibrium results. Through the analysis of the forecast results and the comparison of several error measurement indexes, we verify the rationality of the Gaussian distribution assumption. In addition, an example with three players illustrates the dynamic process of EIM. This example also demonstrates that guideline with penalty term can improve the overall efficiency and forecast accuracy of EIM.

Our work can be extended in many interesting directions. For example, although we have considered the price uncer-

tainties, it is interesting to fully characterize the randomness in the energy procurement price in the formulation. Also, our proposed mechanism can incentivize the information sharing. However, it remains unknown how to design the information sharing market. Especially, to better track the information usage, one promising solution is to implement a blockchain-enabled information sharing market.

APPENDIX A PROOF OF THEOREM 1

Existence of Subgame Perfect Equilibrium:

Note that expected benefit B_i can be expressed as follows,

$$\begin{aligned} \mathbb{E}\{B_i\} &= \mathbb{E}\{\lambda_u \cdot L_i - (\lambda_t \cdot Q_i + \lambda_r \cdot \Delta_i) - C_d(MSE_i)\} \\ &= \lambda_u \cdot L_i + \lambda_t \cdot (\alpha_i - D_i - \mu_i) \\ &\quad - C_d(MSE_i) - \mathbb{E}\{\lambda_r \cdot \Delta_i\}. \end{aligned} \quad (33)$$

As stated above, if there exists a Nash equilibrium, Eq. (13) holds for each player. Since only $\mathbb{E}\{\lambda_r \cdot \Delta_i\}$ and $\lambda_t \cdot \alpha_i$ are related to α_i , the first-order optimality condition reduces to

$$\frac{\partial \mathbb{E}\{\lambda_r \cdot \Delta_i\}}{\partial \alpha_i} = \lambda_t. \quad (34)$$

The only remaining hurdle is to fully characterize $\mathbb{E}\{\lambda_r \cdot \Delta_i\}$ in terms of α_i :

$$\begin{aligned} \mathbb{E}\{\lambda_r \cdot \Delta_i\} &= \alpha_i \cdot \mathbb{E}\{\lambda_r\} - \mathbb{E}\{\lambda_r \cdot e_i\} \\ &= \alpha_i \cdot \{\lambda_s \cdot F_e(\alpha) + \lambda_e \cdot [1 - F_e(\alpha)]\} \\ &\quad - \lambda_s \int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) F_{e|e_i}(e < \alpha | e_i = y_i) dy_i \\ &\quad - \lambda_e \int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) F_{e|e_i}(e > \alpha | e_i = y_i) dy_i \\ &= \alpha_i \cdot \{\lambda_s \cdot F_e(\alpha) + \lambda_e \cdot [1 - F_e(\alpha)]\} - \lambda_e \mu_i \\ &\quad - (\lambda_s - \lambda_e) \int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) F_{e|e_i}(e < \alpha | e_i = y_i) dy_i \\ &= \lambda_e \alpha_i - \lambda_e \mu_i + \alpha_i (\lambda_s - \lambda_e) F_e(\alpha) \\ &\quad - (\lambda_s - \lambda_e) \int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) F_{e|e_i}(e < \alpha | e_i = y_i) dy_i, \end{aligned} \quad (35)$$

where $e_{-i} = \sum_{j=1, j \neq i}^N e_j$, and $\alpha_{-i} = \sum_{j=1, j \neq i}^N \alpha_j$. The functions $f_{e_i}(\cdot)$ and $F_e(\cdot)$ represent probability density function of system total error respectively. In addition, we use $F_{e|e_i}(\cdot)$ and $f_{e|e_i}(\cdot)$ to represent the conditional cumulative distribution function and conditional probability density function between e and e_i respectively.

Note that the following two equations hold:

$$\frac{\partial \alpha_i F_e(\alpha)}{\partial \alpha_i} = F_e(\alpha) + \alpha_i f_e(\alpha), \quad (36)$$

where $f_e(\cdot)$ is probability density function of system total error, and,

$$\begin{aligned} &\frac{\partial \int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) F_{e|e_i}(e < \alpha | e_i = y_i) dy_i}{\partial \alpha_i} \\ &= \int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) f_{e|e_i}(e = \alpha | e_i = y_i) dy_i. \end{aligned} \quad (37)$$

Hence, combining Eq. (36) and (37) with (35) yields

$$\begin{aligned} \frac{\partial \mathbb{E}\{\lambda_r \cdot \Delta_i\}}{\partial \alpha_i} &= \lambda_e + (\lambda_s - \lambda_e) \cdot [F_e(\alpha) + \alpha_i f_e(\alpha)] \\ &- (\lambda_s - \lambda_e) \int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) f_{e|e_i}(e = \alpha | e_i = y_i) dy_i. \end{aligned} \quad (38)$$

According to the definition of conditional expectation, we can find that,

$$\begin{aligned} &\int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) f_{e|e_i}(e = \alpha | e_i = y_i) dy_i \\ &= \int_{-\infty}^{+\infty} y_i f_e(\alpha) f_{e_i|e}(e_i = y_i | e = \alpha) dy_i \\ &= f_e(\alpha) \cdot \mathbb{E}\{e_i | e = \alpha\}, \end{aligned} \quad (39)$$

where $f_{e_i|e}$ is the conditional probability density function between e_i and e .

According to Eq. (34),

$$F_e(\alpha) + f_e(\alpha) (\alpha_i - \mathbb{E}\{e_i | e = \alpha\}) = \frac{\lambda_t - \lambda_e}{\lambda_s - \lambda_e}. \quad (40)$$

Thus, $\alpha_i^* = \mathbb{E}\{e_i | e = \alpha^*\}$ and $F_e(\alpha^*) = \frac{\lambda_t - \lambda_e}{\lambda_s - \lambda_e}$ establish the equilibrium.

Uniqueness of Subgame Perfect Equilibrium:

Summing condition (40) over all players, we know

$$\begin{aligned} LHS &= N \cdot F_e(\alpha) + f_e(\alpha) \left(\sum_{j=1}^N \alpha_j - \sum_{j=1}^N \mathbb{E}\{e_j | e = \alpha\} \right) \\ &= N \cdot F_e(\alpha). \\ RHS &= N \cdot \frac{\lambda_t - \lambda_e}{\lambda_s - \lambda_e}. \end{aligned} \quad (41)$$

Hence, it holds

$$F_e(\alpha) = \frac{\lambda_t - \lambda_e}{\lambda_s - \lambda_e}. \quad (42)$$

Bringing (42) back to Eq. (40) yields that

$$f_e(\alpha) (\alpha_i - \mathbb{E}\{e_i | e = \alpha\}) = 0. \quad (43)$$

Hence, α_i^* is the unique equilibrium if $f_e(\alpha) > 0$. ■

APPENDIX B PROOF OF THEOREM 2

We already know that

$$\begin{aligned} \frac{\partial \mathbb{E}\{B_i\}}{\partial \alpha_i} &= \lambda_t - \lambda_e - (\lambda_s - \lambda_e) \cdot F_e(\alpha) \\ &- (\lambda_s - \lambda_e) f_e(\alpha) \cdot (\alpha_i - \mathbb{E}\{e_i | e = \alpha\}). \end{aligned} \quad (44)$$

For convenience, we define

$$\Psi(\alpha_i) = \lambda_t - \lambda_e - (\lambda_s - \lambda_e) \cdot F_e(\alpha_i + \alpha_{-i}), \quad (45)$$

and

$$\Omega(\alpha_i) = \alpha_i - \mathbb{E}\{e_i | e = \alpha_i + \alpha_{-i}\}. \quad (46)$$

Hence, Eq. (44) can be expressed as

$$\frac{\partial \mathbb{E}\{B_i\}}{\partial \alpha_i} = \Psi(\alpha_i) - (\lambda_s - \lambda_e) f_e(\alpha) \cdot \Omega(\alpha_i). \quad (47)$$

It is obvious that function $\Psi(\alpha_i)$ decreases monotonically with α_i when α_{-i} is determined due to the monotonicity of $F_e(\cdot)$. In addition, $\Psi(\alpha_i^*) = 0$.

Next, we consider the monotonicity of $\Omega(\alpha_i)$. Note that

$$\gamma = \sum_{k=1}^N \mathbb{E}\{e_k | e = \gamma\} \quad (48)$$

is always true. Calculating partial derivative of (48) with respect to γ yields that

$$1 = \sum_{j=1}^N \frac{\partial \mathbb{E}\{e_j | e = \gamma\}}{\partial \gamma}. \quad (49)$$

Combining with Eq. (18), we obtain that,

$$\frac{\partial \mathbb{E}\{e_j | e = \gamma\}}{\partial \gamma} \leq 1. \quad (50)$$

Hence, $\Omega(\alpha_i)$ is monotonically increasing and $\Omega(\alpha_i) = 0$ since

$$\frac{\partial \Omega(\alpha_i)}{\partial \alpha_i} = 1 - \frac{\partial \mathbb{E}\{e_i | e = \alpha_i + \alpha_{-i}\}}{\partial \alpha_i} \geq 0. \quad (51)$$

Together, we find that $\frac{\partial \mathbb{E}\{B_i\}}{\partial \alpha_i}$ decreases monotonically with respect to α_i when α_{-i} is fixed, and $\frac{\partial \mathbb{E}\{B_i\}}{\partial \alpha_i} = 0$ when $\alpha_i = \alpha_i^*$. Hence, when α_{-i} is determined, $\alpha = \alpha_i^*$ is the optimal choice for player i . ■

APPENDIX C PROOF OF THEOREM 3

By transforming the summation of N players' benefit into the form of only one player in the system, the conclusion can be proved. Note that the total expected benefit of all players can be calculated as follows

$$\begin{aligned} &\mathbb{E}\{B\} \\ &= \sum_{j=1}^N \{\lambda_u L_j - (\lambda_t Q_j + \lambda_r \Delta_j) - C_d(MSE_j)\} \\ &= \mathbb{E}\left\{ \lambda_u L - (\lambda_t Q + \lambda_r \Delta) - \sum_{j=1}^N C_d(MSE_j) \right\}, \end{aligned} \quad (52)$$

where $L = \sum_{j=1}^N L_j$ and $Q = \sum_{j=1}^N Q_j$. It is straightforward to verify that

$$\alpha^* = F_e^{-1}\left(\frac{\lambda_t - \lambda_e}{\lambda_s - \lambda_e}\right) \quad (53)$$

is the unique solution to the first-order optimality condition of (52). ■

APPENDIX D
PROOF OF THEOREM 4

We can calculate $\mathbb{E}\{|\Delta|\}$ as follows,

$$\begin{aligned} & \mathbb{E}\{|\Delta|\} \\ &= \int_{-\infty}^{\alpha} f_e(y) \cdot (\alpha - y) dy + \int_{\alpha}^{+\infty} f_e(y) \cdot (y - \alpha) dy \quad (54) \\ &= \alpha - \mu - 2\alpha \int_{\alpha}^{+\infty} f_e(y) dy + 2 \int_{\alpha}^{+\infty} f_e(y) \cdot y dy. \end{aligned}$$

Taking the partial derivative of $\mathbb{E}\{|\Delta|\}$ with respect to α :

$$\begin{aligned} & \frac{\partial \mathbb{E}\{|\Delta|\}}{\partial \alpha} \\ &= 1 - 2 \left(\int_{\alpha}^{+\infty} f_e(y) dy - \alpha f_e(\alpha) \right) - 2\alpha f_e(\alpha) \quad (55) \\ &= 2F_e(\alpha) - 1. \end{aligned}$$

Our theorem immediately follows. \blacksquare

APPENDIX E
PROOF OF THEOREM 5

We know that

$$\begin{aligned} \mathbb{E}\{B_j\} &= \lambda_u L_j - \lambda_t D_j - \lambda_t \mu_j + \lambda_t \alpha_i \\ &\quad - C_d(MSE_i) - \mathbb{E}\{\lambda_r \cdot \Delta_j\}, \quad (56) \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E}\{\lambda_r \cdot \Delta_j\} \\ &= \lambda_e \alpha_j - \lambda_e \mu_j + \alpha_j (\lambda_s - \lambda_e) F_e(\alpha) \quad (57) \\ &\quad - (\lambda_s - \lambda_e) \int_{-\infty}^{+\infty} y_j f_{e_j}(y_j) f_{e|e_j}(e < \alpha | e_j = y_j) dy_j. \end{aligned}$$

Note that the following two equations hold:

$$\frac{\partial F_e(\alpha)}{\partial \alpha_S} = f_e(\alpha), \quad (58)$$

and

$$\begin{aligned} & \frac{\partial \int_{-\infty}^{+\infty} y_j f_{e_j}(y_j) F_{e|e_j}(e < \alpha | e_j = y_j) dy_j}{\partial \alpha_S} \quad (59) \\ &= \int_{-\infty}^{+\infty} y_j f_{e_j}(y_j) f_{e|e_j}(e = \alpha | e_j = y_j) dy_j, \end{aligned}$$

where $\alpha_S = \sum_{i \in S} \alpha_i$.

Hence, combining Eq. (58) and (59) with (57) yields

$$\begin{aligned} & \frac{\partial \mathbb{E}\{\lambda_r \cdot \Delta_j\}}{\partial \alpha_S} = (\lambda_s - \lambda_e) \cdot \alpha_j f_e(\alpha) \quad (60) \\ &\quad - (\lambda_s - \lambda_e) \int_{-\infty}^{+\infty} y_j f_{e_j}(y_j) f_{e|e_j}(e = \alpha | e_j = y_j) dy_j. \end{aligned}$$

According to the definition of conditional expectation, we can find that,

$$\begin{aligned} & \int_{-\infty}^{+\infty} y_i f_{e_j}(y_j) f_{e|e_j}(e = \alpha | e_j = y_j) dy_j \\ &= \int_{-\infty}^{+\infty} y_j f_e(\alpha) f_{e_j|e}(e_j = y_j | e = \alpha) dy_j \quad (61) \\ &= f_e(\alpha) \cdot \mathbb{E}\{e_j | e = \alpha\}. \end{aligned}$$

Thus we have

$$\frac{\partial \mathbb{E}\{B_j\}}{\partial \alpha_S} = (\lambda_e - \lambda_s) f_e(\alpha) \cdot (\alpha_j - \mathbb{E}\{e_j | e = \alpha\}). \quad (62)$$

For convenience, we define

$$\hat{\Omega}(\alpha_S) = \alpha_j - \mathbb{E}\{e_j | e = \alpha_S + \alpha_{-S}\}, \quad (63)$$

where $\alpha_{-S} = \sum_{i \in \{1, 2, \dots, N\} \setminus S}$.

Next, we consider the monotonicity of $\hat{\Omega}(\alpha_i)$. Consider

$$\frac{\partial \hat{\Omega}(\alpha_S)}{\partial \alpha_S} = - \frac{\partial \mathbb{E}\{e_i | e = \alpha_S + \alpha_{-S}\}}{\partial \alpha_S} \leq 0. \quad (64)$$

Hence we have $\frac{\partial \mathbb{E}\{B_j\}}{\partial \alpha_S}$ increases monotonically with respect to α_S when α_{-S} is fixed, and $\frac{\partial \mathbb{E}\{B_j\}}{\partial \alpha_S} = 0$ when $\alpha_i = \alpha_i^*, \forall i \in \{1, 2, \dots, N\}$. Hence, when α_j is fixed to be $\alpha_j^*, \forall j \in \{1, 2, \dots, N\} \setminus S$, $\alpha_S = \alpha_S^*$ minimizes $\mathbb{E}\{B_j\}$. \blacksquare

APPENDIX F
PROOF OF PROPOSITION 1

Before calculating expectation of player i 's benefit function, we introduce two lemmas for later calculations and proofs.

Lemma 1: For Gaussian variable X satisfying $X \sim N(\mu, \sigma^2)$, it holds

$$\int_{\mu}^{+\infty} x \cdot f_X(x) dx = \frac{\sigma}{\sqrt{2\pi}} + \frac{\mu}{2}. \quad (65)$$

Proof: Normalizing $X \sim N(\mu, \sigma^2)$ to be Y ,

$$Y = \frac{X - \mu}{\sigma} \sim N(0, 1). \quad (66)$$

Integrating variable Y from 0 to infinity yields

$$\begin{aligned} & \int_0^{+\infty} y \cdot f_Y(y) dy = \int_0^{+\infty} y \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &= 2 \int_0^{+\infty} \frac{y}{\sqrt{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{y}{\sqrt{2}}\right)^2} d\frac{y}{\sqrt{2}} \quad (67) \\ &= \frac{1}{\sqrt{2\pi}} \Gamma(1) = \frac{1}{\sqrt{2\pi}}, \end{aligned}$$

where the last line of Eq. (67) uses the definition of Gamma function. Generalizing the conclusion to X leads to

$$\begin{aligned} & \int_{\mu}^{+\infty} x \cdot f_X(x) dx \\ &= \int_{\mu}^{+\infty} (\sigma y + \mu) \cdot \frac{1}{\sigma} f_Y(y) d(\sigma y + \mu) \\ &= \int_{\mu}^{+\infty} (\sigma y + \mu) \cdot f_Y(y) dy \quad (68) \\ &= \sigma \cdot \int_{\mu}^{+\infty} y \cdot f_Y(y) dy + \mu \cdot \int_{\mu}^{+\infty} f_Y(y) dy \\ &= \frac{\sigma}{\sqrt{2\pi}} + \frac{\mu}{2}. \end{aligned}$$

Lemma 2: For any variable X and constant a ,

$$\int_0^{+\infty} e^{-ax^2} dx = \frac{1}{2} \sqrt{\frac{\pi}{a}}. \quad (69)$$

Proof: Define Z as follows:

$$Z = \int_{-\infty}^{+\infty} e^{-ax^2} dx. \quad (70)$$

Then, we can first study the behavior of Z^2 ,

$$Z^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-a(x^2+y^2)} dx dy. \quad (71)$$

Replacing (71) with integral of polar coordinates as follows,

$$\begin{aligned} Z^2 &= \int_0^{+\infty} \int_0^{2\pi} e^{-ar^2} r d\theta dr \\ &= -\frac{2\pi}{2a} \cdot \int_0^{+\infty} e^{-ar^2} d(-ar^2) \\ &= -\frac{2\pi}{2a} \cdot e^{-ar^2} \Big|_0^{+\infty} = \frac{\pi}{a}. \end{aligned} \quad (72)$$

The symmetry of Z indicates

$$\int_0^{+\infty} e^{-ax^2} dx = \frac{1}{2} \int_{-\infty}^{+\infty} e^{-ax^2} dx = \frac{1}{2} \sqrt{\frac{\pi}{a}}. \quad (73)$$

Proof for the Property of Expectation:

When $e_j \sim N(0, \sigma_j^2)$ for each player j , then,

$$\alpha_j^* = \mu_j = 0 \quad \forall j \in \{1, \dots, N\}. \quad (74)$$

According to Eqs. (33) and (35), we can obtain,

$$\begin{aligned} \mathbb{E}\{B_i\} &= \lambda_u L_i - \lambda_t D_i - C_0 - \lambda_d c - \lambda_d a MSE_i^{-b} \\ &\quad - (\lambda_s - \lambda_e) \int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) f(e > \alpha | e_i = y_i) dy_i. \end{aligned} \quad (75)$$

Note that

$$\begin{aligned} &\int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) f(e > \alpha | e_i = y_i) dy_i \\ &= \int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) [1 - F_{e_i}(-y_i)] dy_i \\ &= \int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) dy_i - \int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) F_{e_i}(-y_i) dy_i \\ &= - \int_{-\infty}^{+\infty} y_i f_{e_i}(y_i) F_{e_i}(-y_i) dy_i \\ &= \int_0^{+\infty} y_i f_{e_i}(y_i) [F_{e_i}(y_i) - F_{e_i}(-y_i)] dy_i \\ &= \int_0^{+\infty} y_i f_{e_i}(y_i) [2F_{e_i}(y_i) - 1] dy_i \\ &= 2 \int_0^{+\infty} y_i f_{e_i}(y_i) F_{e_i}(y_i) dy_i - \frac{\sigma_i}{\sqrt{2\pi}}, \end{aligned} \quad (76)$$

where the last equation utilizes Lemma 1.

According to integration by parts,

$$\begin{aligned} \int_0^{+\infty} h(y_i) \cdot g'(y_i) dy_i &= h(y_i) g(y_i) \Big|_0^{+\infty} \\ &\quad - \int_0^{+\infty} g(y_i) \cdot h'(y_i) dy_i, \end{aligned} \quad (77)$$

we can calculate the first term in the last line of Eq. (76) by defining $h(y_i)$ and $g(y_i)$ as follows:

$$h(y_i) = F_{e_i}(y_i) = \int_{-\infty}^{y_i} e^{-\frac{x^2}{2\sigma_i^2}} dx, \quad (78)$$

$$g(y_i) = \int y_i f_{e_i}(y_i) dy_i = \int \frac{e^{-\frac{y_i^2}{2\sigma_i^2}}}{2} dy_i^2 = -\sigma_i^2 e^{-\frac{y_i^2}{2\sigma_i^2}}. \quad (79)$$

It holds that

$$\begin{aligned} &\int_0^{+\infty} y_i f_{e_i}(y_i) F_{e_i}(y_i) dy_i \\ &= \frac{1}{\sqrt{2\pi}\sigma_i} \cdot \frac{1}{\sqrt{2\pi}\sigma_{-i}} \int_0^{+\infty} h(y_i) g'(y_i) dy_i \\ &= \frac{-1}{2\pi\sigma_i\sigma_{-i}} \sigma_i^2 e^{-\frac{y_i^2}{2\sigma_i^2}} \cdot \int_{-\infty}^{y_i} e^{-\frac{x^2}{2\sigma_{-i}^2}} dx \Big|_0^{+\infty} \\ &\quad - \frac{1}{2\pi\sigma_i\sigma_{-i}} \int_0^{+\infty} g(y_i) h'(y_i) dy_i \\ &= \frac{\sigma_i}{2\sqrt{2\pi}} + \frac{\sigma_i^2}{2\pi\sigma_i\sigma_{-i}} \int_0^{+\infty} e^{-\frac{y_i^2}{2\sigma_i^2}} e^{-\frac{y_i^2}{2\sigma_{-i}^2}} dy_i \\ &= \frac{\sigma_i}{2\sqrt{2\pi}} + \frac{\sigma_i^2}{2\pi\sigma_i\sigma_{-i}} \int_0^{+\infty} e^{-\frac{y_i^2}{2} \left(\frac{1}{\sigma_i^2} + \frac{1}{\sigma_{-i}^2} \right)} dy_i \\ &= \frac{\sigma_i}{2\sqrt{2\pi}} + \frac{\sigma_i^2}{2\pi\sigma_i\sigma_{-i}} \cdot \frac{\sqrt{\pi}}{2\sqrt{\frac{1}{2} \left(\frac{1}{\sigma_i^2} + \frac{1}{\sigma_{-i}^2} \right)}} \\ &= \frac{\sigma_i}{2\sqrt{2\pi}} + \frac{\sigma_i^2}{2\sqrt{2\pi} \sqrt{\sigma_i^2 + \sigma_{-i}^2}}, \end{aligned} \quad (80)$$

where the fifth equation of Eq. (80) uses Lemma 2 by setting a to be $\frac{1}{2} \left(\frac{1}{\sigma_i^2} + \frac{1}{\sigma_{-i}^2} \right)$. Combining with our observation $\sigma_i^2 = MSE_i$ in (24), our desired conclusion immediately follows.

Proof for the Property of Variance:

To better facilitate the calculation, we define A to be

$$A = \lambda_u L_i - \lambda_t D_i - C_0 - \lambda_d c - \lambda_d a MSE_i^{-b}. \quad (81)$$

The variance of B_i can be calculated as follows,

$$\begin{aligned} D\{B_i\} &= \mathbb{E}\{B_i^2\} - \mathbb{E}\{B_i\}^2 \\ &= A^2 + \lambda_t^2 \mathbb{E}\{e_i^2\} + \mathbb{E}\{\lambda_r^2 e_i^2\} + 2A \mathbb{E}\{\lambda_r e_i\} \\ &\quad - 2\lambda_t \mathbb{E}\{\lambda_r e_i^2\} - A^2 - 2A \mathbb{E}\{\lambda_r e_i\} - \mathbb{E}^2\{\lambda_r e_i\} \\ &= \lambda_t^2 \mathbb{E}\{e_i^2\} + \mathbb{E}\{\lambda_r^2 e_i^2\} - 2\lambda_t \mathbb{E}\{\lambda_r e_i^2\} - \mathbb{E}^2\{\lambda_r e_i\}. \end{aligned} \quad (82)$$

According to Eq. (25), we know

$$\mathbb{E}\{\lambda_r e_i\} = -\mathbb{E}\{\lambda_r \Delta_i\} = \frac{(\lambda_s - \lambda_e) MSE_i}{\sqrt{2\pi} \sqrt{MSE_i + MSE_{-i}}}. \quad (83)$$

Next, we calculate $\mathbb{E}\{e_i^2\}$, $\mathbb{E}\{\lambda_r e_i^2\}$ and $\mathbb{E}\{\lambda_r^2 e_i^2\}$ respectively.

Firstly, for $\mathbb{E}\{e_i^2\}$,

$$\mathbb{E}\{e_i^2\} = D\{e_i\} + \mathbb{E}\{e_i\}^2 = D\{e_i\} + 0 = \sigma_i^2. \quad (84)$$

Secondly, for $\mathbb{E}\{\lambda_r e_i^2\}$,

$$\begin{aligned}
& \mathbb{E}\{\lambda_r e_i^2\} \\
&= (\lambda_s - \lambda_e) \int_{-\infty}^{+\infty} y_i^2 f_{e_i}(y_i) f(e < 0 | e_i = y_i) dy_i \\
&\quad + \lambda_e \int_{-\infty}^{+\infty} y_i^2 f_{e_i}(y_i) dy_i \\
&= (\lambda_s - \lambda_e) \int_0^{+\infty} y_i^2 f_{e_i}(y_i) dy_i + \lambda_e \mathbb{E}\{e_i^2\} \\
&= \frac{\lambda_s - \lambda_e}{2} \int_{-\infty}^{+\infty} y_i^2 f_{e_i}(y_i) dy_i + \lambda_e \mathbb{E}\{e_i^2\} \\
&= \left[\frac{\lambda_s - \lambda_e}{2} + \lambda_e \right] \cdot \mathbb{E}\{e_i^2\} \\
&= \frac{\sigma_i^2 (\lambda_s + \lambda_e)}{2},
\end{aligned} \tag{85}$$

where the second equation of (85) is due to the symmetry of Gaussian distribution, and the fourth equation utilizes Eq. (84).

Lastly, for $\mathbb{E}\{\lambda_r^2 e_i^2\}$,

$$\begin{aligned}
& \mathbb{E}\{\lambda_r^2 e_i^2\} \\
&= \lambda_e^2 \int_{-\infty}^{+\infty} y_i^2 f_{e_i}(y_i) f(e > 0 | e_i = y_i) dy_i \\
&\quad + \lambda_s^2 \int_{-\infty}^{+\infty} y_i^2 f_{e_i}(y_i) f(e < 0 | e_i = y_i) dy_i \\
&= \lambda_e^2 \int_{-\infty}^{+\infty} y_i^2 f_{e_i}(y_i) dy_i \\
&\quad + (\lambda_s^2 - \lambda_e^2) \int_{-\infty}^{+\infty} y_i^2 f_{e_i}(y_i) f(e < 0 | e_i = y_i) dy_i \\
&= \lambda_e^2 \mathbb{E}\{e_i^2\} + (\lambda_s^2 - \lambda_e^2) \frac{\sigma_i^2}{2} \\
&= \frac{\sigma_i^2 (\lambda_s^2 + \lambda_e^2)}{2},
\end{aligned} \tag{86}$$

where the third equation utilizes Eq. (85).

Combining (82), (84), (85) and (86), we now conclude

$$\begin{aligned}
& D\{B_i\} \\
&= \frac{1}{2} \sigma_i^2 \left[(\lambda_e - \lambda_t)^2 + (\lambda_s - \lambda_t)^2 - \frac{(\lambda_e - \lambda_s)^2 \sigma_i^2}{\pi(\sigma_i^2 + \sigma_{-i}^2)} \right] \\
&= \frac{1}{2} MSE_i [(\lambda_e - \lambda_t)^2 + (\lambda_s - \lambda_t)^2] \\
&\quad - \frac{1}{2} MSE_i \cdot \frac{(\lambda_e - \lambda_s)^2 MSE_i}{\pi(MSE_i + MSE_{-i})}.
\end{aligned} \tag{87}$$

APPENDIX G PROOF OF THEOREM 6

We prove monotonicity by solving the partial derivative of $D\{B_i\}$ with respect to MSE_i as follows,

$$\begin{aligned}
& \frac{\partial D\{B_i\}}{\partial MSE_i} \\
&= \frac{(\lambda_e - \lambda_t)^2}{2} + \frac{(\lambda_s - \lambda_t)^2}{2} - \frac{(\lambda_e - \lambda_s)^2 MSE_i}{2\pi(MSE_i + MSE_{-i})} \\
&\quad - \frac{1}{2} MSE_i \cdot \frac{(\lambda_e - \lambda_s)^2}{\pi} \cdot \frac{MSE_{-i}}{(MSE_i + MSE_{-i})^2}.
\end{aligned} \tag{88}$$

The independent zero-mean Gaussian assumption for each forecast error yields that $e \sim N(0, \sum_{j=1}^N \sigma_j^2)$. From (14) and (15) in Theorem 1, we submit

$$\frac{\lambda_t - \lambda_e}{\lambda_s - \lambda_e} = \frac{1}{2}. \tag{89}$$

Hence, it holds

$$(\lambda_e - \lambda_t)^2 = (\lambda_s - \lambda_t)^2 = \frac{1}{4} (\lambda_e - \lambda_s)^2. \tag{90}$$

Eq. (90) can further simplify Eq. (88) as follows

$$\begin{aligned}
& \frac{\partial D\{B_i\}}{\partial MSE_i} \\
&= (\lambda_e - \lambda_t)^2 - \frac{2(\lambda_e - \lambda_t)^2 MSE_i}{\pi(MSE_i + MSE_{-i})} \\
&\quad - \frac{2(\lambda_e - \lambda_t)^2}{\pi} \cdot \frac{MSE_i \cdot MSE_{-i}}{(MSE_i + MSE_{-i})^2} \\
&= (\lambda_e - \lambda_t)^2 \frac{(\pi - 2)(MSE_i + MSE_{-i})^2 + 2MSE_{-i}^2}{\pi(MSE_i + MSE_{-i})^2} \\
&> 0.
\end{aligned} \tag{91}$$

■

APPENDIX H PROOF OF THEOREM 7

The proof of monotonicity of $\mathbb{E}\{B_i\}$ with respect to MSE_{-i} is straightforward. It immediately follows Eq. (25) in Proposition 1.

To prove the monotonicity of $D\{B_i\}$ with respect to MSE_{-i} , we only need to examine the partial derivative of $D\{B_i\}$ with respect to MSE_{-i} as:

$$\frac{\partial D\{B_i\}}{\partial MSE_{-i}} = \frac{MSE_i^2 \cdot (\lambda_e - \lambda_s)^2}{2\pi \cdot (MSE_i + MSE_{-i})^2} > 0. \tag{92}$$

■

APPENDIX I DETAILS OF THE LSTM MODEL

We adopt LSTM to forecast wind power as LSTM can handle the entire sequence of data with high accuracy. In the wind power forecast problem, since wind power is greatly affected by the natural environment, we select five features for the LSTM model: hourly wind speed 2 meters above displacement height, hourly wind speed 10 meters above displacement height, hourly wind speed 50 meters above ground, hourly temperature 2 meters above displacement height and hourly air pressure at surface [29]. When training the LSTM wind power predictor, we set 50 units, 80 epochs, and 100 batch size.

REFERENCES

- [1] A. Kaur, L. Nonnenmacher, H. T. Pedro, and C. F. Coimbra, "Benefits of solar forecasting for energy imbalance markets," *Renewable energy*, vol. 86, pp. 819–830, 2016.
- [2] N. Ghadimi, A. Akbarimajd, H. Shayeghi, and O. Abedinia, "Two stage forecast engine with feature selection technique and improved meta-heuristic algorithm for electricity load forecasting," *Energy*, vol. 161, pp. 130–142, 2018.

- [3] J. Bedi and D. Toshniwal, "Deep learning framework to forecast electricity demand," *Applied Energy*, vol. 238, pp. 1312–1326, 2019.
- [4] Y. Liu, W. Wang, and N. Ghadimi, "Electricity load forecasting by an improved forecast engine for building level consumers," *Energy*, vol. 139, pp. 18–30, 2017.
- [5] W. Qiao and Z. Yang, "Forecast the electricity price of u.s. using a wavelet transform-based hybrid model," *Energy*, vol. 193, p. 116704, 2020.
- [6] H. Chitsaz, P. Zamani-Dehkordi, H. Zareipour, and P. P. Parikh, "Electricity price forecasting for operational scheduling of behind-the-meter storage systems," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6612–6622, 2018.
- [7] A. Bello, D. W. Bunn, J. Reneses, and A. Muñoz, "Medium-term probabilistic forecasting of electricity prices: A hybrid approach," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 334–343, 2017.
- [8] W. Gao, A. Darvishan, M. Toghiani, M. Mohammadi, O. Abedinia, and N. Ghadimi, "Different states of multi-block based forecast engine for price and load prediction," *International Journal of Electrical Power & Energy Systems*, vol. 104, pp. 423–435, 2019.
- [9] M. J. Sanjari and H. B. Gooi, "Probabilistic forecast of pv power generation based on higher order markov chain," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2942–2952, 2017.
- [10] P. Du and J. Matevosyan, "Forecast system inertia condition and its impact to integrate more renewables," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1531–1533, 2018.
- [11] M. Saeedi, M. Moradi, M. Hosseini, A. Emamifar, and N. Ghadimi, "Robust optimization based optimal chiller loading under cooling demand uncertainty," *Applied Thermal Engineering*, vol. 148, pp. 1081–1091, 2019.
- [12] H. Khodaei, M. Hajiali, A. Darvishan, M. Sepehr, and N. Ghadimi, "Fuzzy-based heat and power hub models for cost-emission operation of an industrial consumer using compromise programming," *Applied Thermal Engineering*, vol. 137, pp. 395–405, 2018.
- [13] O. Abedinia, M. Zareinejad, M. H. Doranehgard, G. Fathi, and N. Ghadimi, "Optimal offering and bidding strategies of renewable energy based large consumer using a novel hybrid robust-stochastic approach," *Journal of cleaner production*, vol. 215, pp. 878–889, 2019.
- [14] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, "Using bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 188–201, 2020.
- [15] S. Ben Taieb, R. Huser, R. J. Hyndman, and M. G. Genton, "Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2448–2455, 2016.
- [16] K. Baker, G. Hug, and X. Li, "Energy storage sizing taking into account forecast uncertainties and receding horizon operation," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 1, pp. 331–340, 2017.
- [17] P. Xie, Z. Cai, P. Liu, X. Li, Y. Zhang, and D. Xu, "Microgrid system energy storage capacity optimization considering multiple time scale uncertainty coupling," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5234–5245, 2019.
- [18] T. Zhao, H. Yi, M. Chen, C. Wu, and Y. Xu, "Efficient and robust equilibrium strategies of utilities in day-ahead market with load uncertainty," *arXiv preprint arXiv:1909.05578*, 2019.
- [19] D. Kalathil, C. Wu, K. Poolla, and P. Varaiya, "The sharing economy for the electricity storage," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 556–567, 2019.
- [20] 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), pp. 701–726, Association for Computing Machinery, 2019.
- [21] M. J. Kearns, *The computational complexity of machine learning*. MIT press, 1990.
- [22] R. Doherty and M. O'Malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 587–595, 2005.
- [23] K. Bruninx and E. Delarue, "A statistical description of the error on wind power forecasts for probabilistic reserve sizing," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 3, pp. 995–1002, 2014.
- [24] K. J. Arrow, T. Harris, and J. Marschak, "Optimal inventory policy," *Econometrica: Journal of the Econometric Society*, pp. 250–272, 1951.
- [25] R. Gradwohl and O. Reingold, "Fault tolerance in large games," in *Proceedings of the 9th ACM Conference on Electronic Commerce*, (New York), pp. 274–283, Association for Computing Machinery, 2008.
- [26] H. Yi, M. H. Hajiesmaili, Y. Zhang, M. Chen, and X. Lin, "Impact of the uncertainty of distributed renewable generation on deregulated electricity supply chain," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6183–6193, 2018.
- [27] Netztransparenz., "Wind power data." <https://www.netztransparenz.de/>, Accessed on Oct. 12, 2020.
- [28] SMARD., "Consumption data." <https://www.smard.de>, Accessed on Oct. 21, 2020.
- [29] Open Power System Data., "Weather data." <https://open-power-system-data.org/>, Accessed on Oct. 13, 2020.
- [30] Clean Energy Wire., "Price data." <https://www.cleaneenergywire.org/factsheets/what-german-households-pay-power/>, Accessed on Oct. 20, 2020.
- [31] C. Kost, J. N. Mayer, J. Thomsen, N. Hartmann, C. Senkpiel, S. Philipps, S. Nold, S. Lude, N. Saad, and T. Schlegl, "Levelized cost of electricity renewable energy technologies," *Fraunhofer Institute for Solar Energy Systems ISE*, vol. 144, 2013.
- [32] P. Friederichs and T. L. Thorarindottir, "Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction," *Environmetrics*, vol. 23, no. 7, pp. 579–594, 2012.



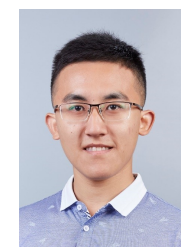
Jingshi Cui (Graduate Student Member, IEEE) is a PhD student at IIS, Tsinghua University, advised by Professor Chenye Wu. Ms. Cui received her bachelor's degree from the Department of Computer Science and Technology, Harbin Institute of Technology, Weihai. She has been awarded National Scholarships in 2016 and 2017, respectively.

Her current research interests include economic analysis and market design in the power system.



Nan Gu (Graduate Student Member, IEEE) is a PhD student at IIS, Tsinghua University, advised by Professor Chenye Wu. Ms. Gu received her bachelor's degree from the Department of Electrical Engineering, Tsinghua University. She has been awarded Excellent Graduate of Tsinghua University in 2019.

Her research interests include market design and optimization in the power system.



Tianyu Zhao received the B.Eng. degree from the School of Energy and Power Engineering and the Minor's Diploma in Business Administration from the School of Management, Xi'an Jiaotong University, Xi'an, China, in 2017. He is currently working toward the Ph.D. degree with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. His research interests include electricity market, microgrid operation, and machine learning applications in power systems.



Chenye Wu (Member, IEEE) is an Assistant Professor and the Presidential Young Fellow at School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen. Before joining CUHK Shenzhen, he was an Assistant Professor at IIS, Tsinghua University. He worked at ETH Zurich as a Research Scientist, working with Professor Gabriela Hug, in 2016. Before that, Professor Kameshwar Poolla and Professor Pravin Varaiya hosted Dr. Wu as a postdoctoral researcher at UC Berkeley for two years. In 2013-2014, He spent one year at Carnegie

Mellon University as a postdoc fellow, hosted by Professor Gabriela Hug and Professor Soumya Kar.

Dr. Wu received his Ph.D. from IIS, Tsinghua University, in July 2013. His Ph.D. advisor is Professor Andrew Yao, the laureate of the A.M. Turing Award in the year of 2000. Dr. Wu was the best paper award co-recipients of IEEE SmartGridComm 2012, IEEE PES General Meeting 2013, IEEE PES General Meeting 2020. Currently, he is working on economic analysis, optimal control and operation of power systems.



Minghua Chen (S'04 M'06 SM' 13) (Senior Member, IEEE) received his B.Eng. and M.S. degrees from the Department of Electronic Engineering at Tsinghua University. He received his Ph.D. degree from the Department of Electrical Engineering and Computer Sciences at University of California Berkeley. He was with Microsoft Research Redmond and The Chinese University of Hong Kong before joining School of Data Science, City University of Hong Kong, where he is a Professor. Minghua received the Eli Jury award from UC Berkeley in

2007 (presented to a graduate student or recent alumnus for outstanding achievement in the area of Systems, Communications, Control, or Signal Processing) and The Chinese University of Hong Kong Young Researcher Award in 2013. He also received several best paper awards, including IEEE ICME Best Paper Award in 2009, IEEE Transactions on Multimedia Prize Paper Award in 2009, ACM Multimedia Best Paper Award in 2012, and IEEE INFOCOM Best Poster Award in 2021. He serves as Associate Editor of IEEE/ACM Transactions on Networking in 2014 - 2018. He serves as TPC Co-Chair, General Chair, and Steering Committee Chair of ACM e-Energy in 2016, 2017, and 2018 - present, respectively. He serves as TPC Co-Chair for ACM MobiHoc 2020. He receives the ACM Recognition of Service Award in 2017 for the service contribution to the research community. He is currently a Senior Editor for IEEE Systems Journal (2021- present), an Area Editor of ACM SIGEnergy Energy Informatics Review (2021 - present), and an Executive Committee member of ACM SIGEnergy (2018 - present). Minghua's recent research interests include online optimization and algorithms, machine learning in power system operations, intelligent transportation systems, distributed optimization, delay-constrained network coding, and capitalizing the benefit of data-driven prediction in algorithm/system design. He is a Distinguished Member of ACM.