

The following publication Pan, K., & Guan, Y. (2017). Data-driven risk-averse stochastic self-scheduling for combined-cycle units. IEEE Transactions on Industrial Informatics, 13(6), 3058-3069 is available at <https://doi.org/10.1109/TII.2017.2710357>.

Data-Driven Risk-Averse Stochastic Self-Scheduling for Combined-Cycle Units

Kai Pan, *Member, IEEE*, and Yongpei Guan, *Senior Member, IEEE*

Abstract—With fewer emissions, higher efficiency, and quicker response than traditional coal-fired thermal power plants, the combined-cycle units, as gas-fired generators, have been increasingly adapted in the U.S. power system to enhance the smart grids operations. Meanwhile, due to the inherent uncertainties in the deregulated electricity market, e.g., intermittent renewable energy output, unexpected outages of generators and transmissions, and fluctuating electricity demands, the electricity price is volatile. As a result, this brings challenges for an independent power producer (served in the self-scheduling mode) owning combined-cycle units, to maximize the total profit when facing the significant price uncertainties. In this paper, a data-driven risk-averse stochastic self-scheduling approach is presented for the combined-cycle units that participate in the real-time market. The proposed approach does not require the specific distribution of the uncertain real-time price. Instead, a confidence set for the unknown distribution is constructed based on the historical data. The conservatism of the proposed approach is adjustable based on the amount of available data. Finally, numerical studies show the effectiveness of the proposed approach.

Index Terms—Data-driven, stochastic optimization, combined-cycle units, self-scheduling.

I. NOMENCLATURE

A. Sets

\mathcal{H}	Set of all edges in the state transition graph.
$\mathcal{H}_m^{\text{all}}$	Set of all edges connected to mode m .
$\mathcal{H}_m^{\text{in}}$	Set of incoming edges of mode m .
$\mathcal{H}_m^{\text{out}}$	Set of outgoing edges of mode m .
$\mathcal{H}_m^{\text{sl}}$	Set of self-loop edges of mode m .
$\mathcal{H}_i^{\text{sd}}$	Set of edges where turbine i shuts down.
$\mathcal{H}_i^{\text{su}}$	Set of edges where turbine i starts up.
\mathcal{G}^{CT}	Set of CTs.
\mathcal{G}^{ST}	Set of STs.
\mathcal{M}	Set of modes in the state transition graph.
$\mathcal{M}_i^{\text{off}}$	Set of modes where turbine i is offline.
$\mathcal{M}_i^{\text{on}}$	Set of modes where turbine i is online.

B. Parameters

T	Number of time periods in planning horizon, with each period to be one hour.
\bar{C}_m	Generation upper bound of mode m (MW).
\underline{C}_m	Generation lower bound of mode m (MW).

This research was partially supported by the US National Science Foundation under grant ECCS1609794. The work of K. Pan was partially supported by Hong Kong Polytechnic University under grants 1-ZE73 and G-UABE.

Kai Pan is with the Department of Logistics and Maritime Studies, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. E-mail: kai.pan@polyu.edu.hk.

Yongpei Guan is with the Department of Industrial and Systems Engineering, University of Florida, Gainesville, Florida 32611, USA. E-mail: guan@ise.ufl.edu.

Copyright (c) 2009 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

\hat{C}	Total capacity of the combined-cycle unit (MW).
V_e^+	Ramp-up rate limit at edge e (MW/h).
V_e^-	Ramp-down rate limit at edge e (MW/h).
SU_i	Start-up cost of turbine i (\$).
SD_i	Shut-down cost of turbine i (\$).
T_i^{cold}	Cold start time of turbine i (h).
T_i^{warm}	Warm start time of turbine i (h).
T_i^{mu}	Minimum-up time for turbine i (h).
T_i^{md}	Minimum-down time for turbine i (h).
$q_t(\xi)$	Electricity price at time period t corresponding to scenario ξ (\$/MWh).
C. First-Stage Decision Variables	
y_e^t	Binary variable to indicate the status of edge e in the transition graph at time period t , “1” if edge e is active at t and “0” otherwise.
u_i^t	Start-up cost of turbine i at time period t (\$).
v_i^t	Shut-down cost of turbine i at time period t (\$).
D. Second-Stage Decision Variables	
$\phi_m^t(\xi)$	Generation cost of mode m at time period t corresponding to scenario ξ (\$).
$x_m^t(\xi)$	Power generation amount of mode m at time period t corresponding to scenario ξ (MW).

II. INTRODUCTION

This paper considers the self-scheduling problem that an independent power producer (IPP) faces when participating in the deregulated real-time electricity market, where an independent system operator (ISO) takes real-time generation offers and clears the real-time market by considering the load discrepancy between day-ahead and real-time markets [1]. To maximize its own profit by participating in the real-time market, besides the traditional three-part offer approach [2], each IPP can also use “self-commitment” [3] and “self-scheduling” modes [4], [5]. For the “self-commitment” mode, an IPP decides the unit commitment (i.e., generator online/offline status) and submit it to the ISO, who will decide the generation amount for each time period. For the “self-scheduling” mode, an IPP makes the decisions on both the unit commitment and power generation amount before submitting the selling bids to the ISO. Thus, the difference between these two modes is on who decides the generation amount. For the “self-scheduling” mode, a fundamental task for an IPP is to self-schedule its generators so that the total profit obtained from selling energy to the electricity market is maximized [6]. For extensive works on the self-scheduling problem, the readers are referred to [7], [8], and [9], among which the thermal and hydro units are studied in detail. This paper considers the combined-cycle unit (CCU) that received little attention on the self-scheduling problem, although its self-commitment mode has been recently

studied in [3]. Accordingly, an IPP is served as a price-taker and an optimal strategy is derived for a CCU.

Due to the advantages of their physical characteristics, such as quick and flexible responses, low emissions, and the affordable gas price, among others, the CCUs are becoming more and more popular in U.S. to support smart grids operations. It is appealing that CCUs are being widely deployed by major ISOs of the U.S., e.g., MISO [10], PJM [11], ERCOT [12], etc. A CCU consists of several combustion turbines (CTs) and steam turbines (STs), and it can work at different modes with each mode including a part of CTs and STs ([13] and [14]). With the given electricity prices, the self-scheduling decision of a CCU includes the hourly on/off status of each individual turbine (CTs and STs) and the power generation amount of each mode.

On the other hand, with the increasing penetration of renewable energy, unexpected generator and transmission outages, and fluctuating electricity demands, the electricity prices are significantly volatile, which provides a chance for the flexible CCU to gain more profit by participating in the real-time market using the “self-scheduling” mode. However, this uncertainty brings significant challenges to an IPP for making an optimal self-scheduling decision under price uncertainty.

Stochastic optimization approach has been studied extensively to deal with these challenges, as it fits well for the current deregulated electricity market. For instance, a stochastic mixed-integer optimization formulation accommodating price uncertainty is established in [15] by utilizing an estimation of price probability density functions. In [16], a stochastic self-scheduling model is formulated for an IPP participating in three markets including the day-ahead market, the automatic generation control market, and the real-time market, with price volatility represented by a scenario tree. In addition, another scenario tree-based stochastic optimization model is established in [5], where hourly uncertain electricity prices are considered and risk constraints are incorporated. Furthermore, in [3], a two-stage stochastic optimization framework is adopted to make self-commitment decisions for the CCUs considering real-time market price uncertainty and the influence of risk aversion.

Robust optimization approach is another approach to deal with price uncertainty and has been explored in much literature on power system operations since it can ensure the system feasibility and reliability under the worst-case scenario. Early works demonstrating the effectiveness of robust optimization for the unit commitment (UC) problems appear in [17], [18], and [19], in which the security-constrained UC problem is modeled and solved through robust optimization techniques. For the self-scheduling problem under price uncertainty, robust optimization also plays an important role since it is distribution-free [20]. That is, it does not require the precise distribution of the electricity prices, as the stochastic optimization approach does. Along this direction, a robust mixed-integer linear programming model with min-max cost criterion is proposed in [20] for a price-taker power producer who participates in the pool-based electricity market. Price uncertainties are represented with confidence intervals, which are successively divided into a sequence of nested subintervals

and enables the problem to be easily solved.

However, both stochastic and robust optimization approaches face challenges in practice. As mentioned above, for the stochastic self-scheduling model, which uses the stochastic optimization approach, the electricity prices are usually assumed to follow a certain distribution through price forecast and estimation. Then, a sampling approach is conducted to generate a finite number of scenarios that could be realized in the future, with a certain probability corresponding to each scenario. This approach, however, leads to a difficult large-sized stochastic programming model, which is extremely computationally expensive. In contrast, the robust optimization approach is distribution-free. Instead of requiring the detailed distribution of electricity prices, limited information is needed to construct the confidence set of the uncertain prices, which could help reduce the computational time. However, the robust optimization approach always considers the worst-case scenario in the system. This leads to a conservative solution, because the worst-case scenario rarely happens in practice. Therefore, although robust optimization requires less information to represent the price uncertainty, it can be too conservative due to its objective of minimizing the worst-case cost.

Currently, a large amount of historical data on electricity prices are available to public for each ISO/RTO online. Without explicit price distribution, through observing the historical data, a confidence set for the unknown price distribution can be constructed and the self-scheduling model can be further established. Meanwhile, it is worthwhile exploring how the price data would affect the self-scheduling decisions and total profits. Therefore, a data-driven stochastic optimization model considering price uncertainty is proposed in this paper to solve the CCU self-scheduling problem. Historical data are observed to construct the confidence set of price distribution (instead of the price uncertainty set), and different norms are applied to measure the distance between the empirical price distribution and the true price distribution. The proposed model is solved with the objective of maximizing the total expected profit under the worst-case distribution, leading to a risk-averse solution. For this data-driven risk-averse stochastic optimization model, as more data are observed, the distance between the empirical distribution and the true distribution becomes smaller and smaller, i.e., the empirical distribution converges to the true distribution based on nonparametric statistical analysis. As a result, the proposed model converges to the risk-neutral stochastic optimization model. The proposed risk-averse model is also related to risk based approaches. There has been extensive literature on applying risk-averse optimization approach to deal with the problems in smart grid operations. For instance, a risk-averse stochastic UC model is introduced in [21] to hedge against the risk from the load, wind, and photovoltaic (PV) generation uncertainties. In [22], a risk-averse optimization model is constructed to maximize the profits of the plug-in electric vehicle (PEV) under uncertain market prices and fleet mobility. Moreover, risk neutral and risk averse options are compared in [23] when deciding the optimal day-ahead schedule under uncertainty by developing stochastic optimization models for the micro grid

operations. Different risk measures such as value-at-risk [24] and conditional value-at-risk [25], [26] can be applied.

Although the data-driven approach has been applied in fault detection [27], system reliability [28], quality prediction [29], and system prognostic [30], it was rarely applied in the smart grid operations. By introducing the data-driven concept to smart grid operations, the proposed approach in this paper can help accommodate the disadvantages of both stochastic and robust optimization approaches by utilizing the historical data and then provide better decision-marking. The main contributions are described as follows:

- (1) A self-scheduling approach is introduced for IPPs to submit an offer for a combined-cycle unit, for which IPPs have better controls of their assets, as compared to the three-part offering and the self-commitment approaches.
- (2) The data-driven stochastic optimization model does not require the precise price distribution and can be solved directly based on the observed historical data.
- (3) The conservatism of the data-driven risk-averse stochastic optimization model can be adjusted based on how much data are observed, since the confidence set of price distribution is adjustable based on the amount of data with a given level of confidence guarantee. Finally, the conservatism of the proposed model vanishes as the amount of available data increases to infinity. The value of data is also provided by showing how the objective value changes based on the change of historical data.

The remaining part of this paper is organized as follows. The nominal formulation for CCU self-scheduling is first reported in Section III. Then, in Section IV the data-driven stochastic self-scheduling model for a CCU is described. In Section V, a solution approach based on Bender's decomposition framework is elaborated, and the corresponding computational results are presented in Section VI. Finally, this paper is summarized in Section VII.

III. NOMINAL FORMULATION FOR CCU SELF-SCHEDULING

This section describes the nominal formulation of the deterministic CCU self-scheduling model based on the edge-based model described in [31]. For a CCU with m CTs and n STs, a mode in which this unit can work can be represented as a CTs+ b STs, with $0 \leq a \leq m$ and $0 \leq b \leq n$. Note that the steam turbine cannot start up unless the corresponding combustion turbine is online, and thus $b \leq a$. In addition, the CTs are distinguished as CT_1, CT_2, \dots, CT_m and the STs as ST_1, ST_2, \dots, ST_n .

As shown in Figure 1, there are transition edges indicating the transitions between different modes. (As opposed to the configuration-based model [14], this paper focuses on the edges connecting different modes to describe exactly the physical characteristics of each CT and ST in a CCU [31].) For each mode i , as shown in Figure 1, there are incoming edges such as E_{mi} and E_{ni} indicating that modes m and n can transit to mode i . Meanwhile, mode i can stay on the same mode on two consecutive time periods through the self-loop edge E_{ii} or transition to other modes, e.g., mode j and

mode k , through outgoing edges E_{ij} and E_{ik} . For illustrative purposes, the transition edges of the state transition paragraph for a CCU with 2CTs+1ST are summarized in Table I.

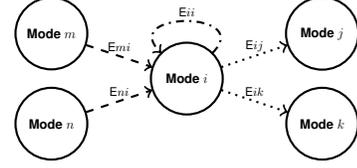


Fig. 1. An example for mode transitions

TABLE I
TRANSITION EDGES FOR 2CTs+1ST

Mode No	m CTs+ n STs	Self-loop edge	Incoming edges	Outgoing edges
0	0CT+0ST	E00	E10, E20, E50	E01, E02, E05
1	CT ₁	E11	E01, E31, E51	E10, E13, E15
2	CT ₂	E22	E02, E42, E52	E20, E24, E25
3	CT ₁ +ST	E33	E13, E63	E31, E36
4	CT ₂ +ST	E44	E24, E64	E42, E46
5	CT ₁ +CT ₂	E55	E05, E15, E25, E65	E50, E51, E52, E56
6	CT ₁ +CT ₂ +ST	E66	E36, E46, E56	E63, E64, E65

Now, the deterministic self-scheduling model is provided for a single CCU to maximize the total profit, i.e., the total revenue minus the total cost, as follows:

$$\max \sum_{t=1}^T \left(\sum_{m \in \mathcal{M}} (q_t x_m^t - \phi_m^t) - \sum_{i \in \mathcal{G}^{\text{CT}}} (u_i^t + v_i^t) \right)$$

$$s.t. \sum_{e \in \mathcal{H}} y_e^t = 1, \forall t, \quad (1)$$

$$\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t = \sum_{e \in (\mathcal{H}_m^{\text{out}} \cup \mathcal{H}_m^{\text{sl}})} y_e^{t+1}, \forall m \in \mathcal{M}, \forall t, \quad (2)$$

$$\sum_{e \in \mathcal{U}_{m \in \mathcal{M}_i^{\text{off}}} \mathcal{H}_m^{\text{all}}} y_e^\tau \leq 1 - \sum_{e \in \mathcal{H}_i^{\text{su}}} y_e^t, \forall i \in \mathcal{G}^{\text{CT}} \cup \mathcal{G}^{\text{ST}}, \quad (3)$$

$$\sum_{e \in \mathcal{U}_{m \in \mathcal{M}_i^{\text{on}}} \mathcal{H}_m^{\text{all}}} y_e^\tau \leq 1 - \sum_{e \in \mathcal{H}_i^{\text{sd}}} y_e^t, \forall i \in \mathcal{G}^{\text{CT}} \cup \mathcal{G}^{\text{ST}}, \quad (4)$$

$$\tau \in \{t+1, \dots, \min\{T, T_i^{\text{md}} + t - 1\}\}, \forall t, \quad (5)$$

$$u_i^t \geq \text{SU}_i^{\text{hot}} \sum_{e \in \mathcal{H}_i^{\text{su}}} y_e^t, \forall i \in \mathcal{G}^{\text{CT}}, \forall t, \quad (6)$$

$$u_i^t \geq \text{SU}_i^{\text{warm}} \left(\sum_{e \in \mathcal{H}_i^{\text{su}}} y_e^t - \sum_{\tau=T_i^{\text{md}}+1}^{T_i^{\text{warm}}} \sum_{e \in \mathcal{H}_i^{\text{sd}}} y_e^{t-\tau} \right), \forall i \in \mathcal{G}^{\text{CT}}, \forall t, \quad (7)$$

$$u_i^t \geq \text{SU}_i^{\text{cold}} \left(\sum_{e \in \mathcal{H}_i^{\text{su}}} y_e^t - \sum_{\tau=T_i^{\text{md}}+1}^{T_i^{\text{cold}}} \sum_{e \in \mathcal{H}_i^{\text{sd}}} y_e^{t-\tau} \right), \forall i \in \mathcal{G}^{\text{CT}}, \forall t, \quad (8)$$

$$v_i^t = \text{SD}_i \sum_{e \in \mathcal{H}_i^{\text{sd}}} y_e^t, \forall i \in \mathcal{G}^{\text{CT}}, \forall t, \quad (9)$$

$$\phi_m^t \geq \alpha_m^n \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right) + \beta_m^n x_m^t, \quad m \in \mathcal{M}, \forall t, n = 1, \dots, N, \quad (10)$$

$$\underline{C}_m \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right) \leq x_m^t, \forall m \in \mathcal{M}, \forall t, \quad (11)$$

$$x_m^t \leq \bar{C}_m \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right), \forall m \in \mathcal{M}, \forall t, \quad (11)$$

$$\sum_{m \in \mathcal{M}} x_m^{t+1} - \sum_{m \in \mathcal{M}} x_m^t \leq V_e^+ y_e^{t+1} + \hat{C}(1 - y_e^{t+1}), \quad \forall e \in \mathcal{H}, \forall t, \quad (12)$$

$$\sum_{m \in \mathcal{M}} x_m^t - \sum_{m \in \mathcal{M}} x_m^{t+1} \leq V_e^- y_e^{t+1} + \hat{C}(1 - y_e^{t+1}), \quad \forall e \in \mathcal{H}, \forall t, \quad (13)$$

$$y_e^t \in \{0, 1\}, x_m^t \geq 0, \phi_m^t \text{ free}, \forall m \in \mathcal{M}, \forall e \in \mathcal{H}, \forall t. \quad (14)$$

In the above formulation, $\sum_{i \in \mathcal{G}^{\text{CT}}} (u_i^t + v_i^t)$ represents the total transition costs (usually the summation of start-up and shut-down costs for all CTs) at time period t . Constraints (1) indicate that only one of all edges is active at each time period since the CCU moves along one edge at each time period.

Constraints (2) indicate the flow balance constraint. That is, for each mode m , the summation of the incoming edge flow at time t is equal to the summation of the outgoing edge flow at time $t+1$. For instance, for mode 0 in Table I, $y_{E00}^t + y_{E10}^t + y_{E20}^t + y_{E50}^t = y_{E00}^{t+1} + y_{E01}^{t+1} + y_{E02}^{t+1} + y_{E05}^{t+1}, \forall t$.

Constraints (3) represent the minimum-up time limits. For each turbine in CCU, once it starts up, the modes that do not contain this turbine to be online should be offline for the following T_i^{mu} consecutive time periods (or $T - t + 1$ time periods if $T < T_i^{\text{mu}} + t - 1$). For instance, for CT_2 in Table I with $T_{\text{CT}_2}^{\text{mu}} = 2$ and $t + 1 \leq T$, the edges indicating that CT_2 starts up (i.e., $\mathcal{H}_{\text{CT}_2}^{\text{su}}$) are E02, E05, E15, and E36; the modes that do not contain CT_2 are Modes 0, 1, and 3. Thus it follows that $y_{E00}^{t+1} + y_{E10}^{t+1} + y_{E20}^{t+1} + y_{E50}^{t+1} + y_{E01}^{t+1} + y_{E02}^{t+1} + y_{E05}^{t+1} + y_{E11}^{t+1} + y_{E31}^{t+1} + y_{E51}^{t+1} + y_{E13}^{t+1} + y_{E15}^{t+1} + y_{E33}^{t+1} + y_{E63}^{t+1} + y_{E36}^{t+1} \leq 1 - y_{E02}^t - y_{E05}^t - y_{E15}^t - y_{E36}^t$, where the left-hand side is the summation of the statuses of the edges connected to Modes 0, 1, and 3, since $\bigcup_{m \in \mathcal{M}_i^{\text{off}}} \mathcal{H}_m^{\text{all}}$ is the set of edges connected to the modes with turbine i offline.

Similarly, constraints (4) represent the minimum-down time limits and show that if turbine i shuts down at t , then the modes that contain this turbine should be offline for the following consecutive minimum-down time periods. Note here that the term $\bigcup_{m \in \mathcal{M}_i^{\text{on}}} \mathcal{H}_m^{\text{all}}$ is the set of the edges connected to the modes with turbine i online.

Constraints (5) - (8) represent the transition cost constraints, with $\text{SU}_i^{\text{hot}}, \text{SU}_i^{\text{warm}}, \text{SU}_i^{\text{cold}}$, and SD_i representing the hot start-up, warm start-up, cold start-up, and shut-down costs, respectively, and $\text{SU}_i^{\text{hot}} < \text{SU}_i^{\text{warm}} < \text{SU}_i^{\text{cold}}$. The transition among different modes (i.e., one of the incoming and outgoing edges is active) results in transition cost for the CCU. Denote it as the summation of start-up and shut-down costs involved in this transition. Note that no transition cost is triggered for each self-loop edge. In addition, the start-up/shut-down costs for STs are omitted here due to the fact that ST is driven by CTs. For each CT, u_i^t is calculated through constraints (5) - (7) by considering three types of start-up costs (i.e., cold start-up cost, warm start-up cost, and hot start-up cost). The calculation of v_i^t is given by constraints (8). For more details on explaining constraints (5) - (7), readers are referred to [31].

Constraints (9) show how to calculate the generation cost that is normally a quadratic curve [32], i.e., $a(x_m^t)^2 +$

$b x_m^t + c \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right)$, which can be represented by an N -piece piecewise linear approximation. Note here that for a CCU, the generation cost at each time period is equal to the total generation costs of all modes at each time period, since there is only one mode online for each time period. Parameters α and β are defined in the piecewise linear approximation. Here $\left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right)$ represents the status of mode m at t , since mode m online at t means one of the incoming or self-loop edges of mode m is active at t .

Constraints (10) - (11) represent the generation amount limits. Similar to the generation cost, the generation amount of a CCU is the summation of the generation amounts over all the modes.

Constraints (12) - (13) represent the ramping constraints, where the right sides indicate that if an edge e is not active (i.e., $y_e^{t+1} = 0$), then the ramping constraints are relaxed, following the definition of \hat{C} . Otherwise, if this edge e is active (i.e., $y_e^{t+1} = 1$), then this edge provides the ramp rate limit for the whole CCU. Note here that constraints (12) and (13) describe the ramp rate limit for both self-loop edges and incoming/outgoing edges, with the former representing the ramp rate within a mode and the latter representing that of the transition between different modes.

IV. DATA-DRIVEN STOCHASTIC SELF-SCHEDULING MODEL

In this section, a two-stage risk-averse stochastic optimization model for the combined-cycle unit self-scheduling problem is firstly proposed and then the confidence set construction for the unknown price distribution and the corresponding convergence analysis are discussed.

A. Two-Stage Risk-Averse Stochastic Optimization Model

The proposed model is extended from the traditional two-stage stochastic optimization framework, and the distribution of random parameter is denoted as \mathbb{P} . In the proposed model, the decisions of edge statuses (i.e., binary variable y) are made in the first stage, and power generation amounts (i.e., continuous variable x) are made in the second stage under the worst-case price distribution, as the objective is to maximize the total expected profit under the worst-case price distribution. For the traditional two-stage stochastic optimization model, \mathbb{P} is given, which means the probability corresponding to each possible scenario ξ is assumed to be known. However, as it is biased to assume any particular distribution for \mathbb{P} , in this paper, the distribution \mathbb{P} is set to be unknown and it is assumed ambiguous (denoted as P for notation clarification). Instead, historical data is available, and based on a given set of historical data, a pre-defined confidence set \mathcal{A} for P is constructed so that the probability distribution P can run freely within set \mathcal{A} . The detailed description on the construction of set \mathcal{A} is provided in Section IV-B. Accordingly, the detailed formulation of the proposed model is described as follows:

$$\max - \sum_{t=1}^T \sum_{i \in \mathcal{G}^{\text{CT}}} (u_i^t + v_i^t) + \min_{P \in \mathcal{A}} \mathbb{E}_P [Q(y, \xi)] \quad (15)$$

s.t. (1) – (8), $y_e^t \in \{0, 1\}, \forall e \in \mathcal{H}, \forall t$,

where $\mathbb{E}_P[\cdot]$ means the expectation under P and $Q(y, \xi)$ is equal to

$$\begin{aligned} \max \quad & \sum_{t=1}^T \sum_{m \in \mathcal{M}} (q_t(\xi) x_m^t(\xi) - \phi_m^t(\xi)) \\ \text{s.t.} \quad & (9) - (13), \forall \xi; \\ & \phi_m^t(\xi) \text{ free}, x_m^t(\xi) \geq 0, \forall m \in \mathcal{M}, \forall t. \end{aligned} \quad (16)$$

This study assumes that ξ has a finite support, i.e., ξ_1, ξ_2, \dots , and ξ_J , indicating a finite number of possible realizations. Then it follows that

$$\mathbb{E}_P[Q(y, \xi)] = \sum_{j=1}^J p_j Q(y, \xi^j), \quad (17)$$

where p_j is the probability corresponding to scenario j . Since P is ambiguous, the probability $p_j, j = 1, \dots, J$, for each scenario is random and satisfies the constraints describing \mathcal{A} in the model.

B. Confidence Set Construction and Convergence Analysis

Now describe how to construct \mathcal{A} based on historical data. Given a set of historical data, e.g., S samples, J bins are constructed to separate the data so that each bin consists of S_1, S_2, \dots, S_J samples. In this way, a histogram with $S = \sum_{j=1}^J S_j$ are constructed. Consequently, the empirical distribution for the uncertain parameter based on the historical data is described $P_0^\top = (p_1^0, p_2^0, \dots, p_J^0)$, with $p_1^0 = S_1/S, p_2^0 = S_2/S, \dots$, and $p_J^0 = S_J/S$. For better illustration purpose, a simple example is given as follows. Given $S = 10$ one-dimensional data samples corresponding to an uncertain parameter, i.e., 5, 12, 16, 21, 30, 43, 55, 72, 89, and 93, five bins can be constructed as Bin 1: $(-\infty, 20)$, Bin 2: $[20, 40)$, Bin 3: $[40, 60)$, Bin 4: $[60, 80)$, and Bin 5: $[80, \infty)$. Through separating the data samples into each bin, it follows that the number of data samples falling into each bin is $S_1 = 3, S_2 = 2, S_3 = 2, S_4 = 1$, and $S_5 = 2$, respectively. Thus, the empirical distribution for the uncertain parameter is described as $P_0^\top = (p_1^0, p_2^0, p_3^0, p_4^0, p_5^0) = (3/10, 2/10, 2/10, 1/10, 2/10)$.

Since the true distribution might be different from the empirical distribution, statistical inference is used to define the confidence set \mathcal{A} , which includes the possible realizations of the true distribution P . Based on the given set of historical data, the distance (denoted as θ) between the true distribution and the empirical distribution can be estimated under different distance measures (or metrics). For better explaining the concept of distance measure, a simple example is given as follows. In the real space \mathbb{R}^n , the distance between two points (vectors) such as $\mathbf{r}^1 = (r_1^1, r_2^1, \dots, r_n^1)$ and $\mathbf{r}^2 = (r_1^2, r_2^2, \dots, r_n^2)$ can be represented by Euclidean norm (or L_2 norm) as $d_2(\mathbf{r}^1, \mathbf{r}^2) = \sqrt{\sum_{i=1}^n (r_i^1 - r_i^2)^2}$. Note here that the distance between \mathbf{r}^1 and \mathbf{r}^2 can also be represented by other distance measures such as L_1 norm.

In this paper, two norms, L_1 and L_∞ norms, are applied to measure the distance between the true distribution P and

the empirical distribution P_0 and thus to construct two types of confidence sets. These two norms are utilized since the empirical distribution converges to the true distribution under these two norms as the amount of available historical data (i.e., S) goes to infinity. That is, these two norms have advantages in guaranteeing the convergence. On the other hand, with these two norms, the model can be reformulated as a mixed-integer linear programming (MILP) formulation, which can be solved easily by commercial optimization solvers (e.g., CPLEX) and then better applied in practices.

First, the confidence sets, \mathcal{A}_1 and \mathcal{A}_∞ , are defined corresponding to L_1 and L_∞ norms, respectively: $\mathcal{A}_1 = \{P \in \mathbb{R}_+^J \mid d_1(P, P_0) = \|P - P_0\|_1 \leq \theta\} = \{P \in \mathbb{R}_+^J \mid \sum_{j=1}^J |p_j - p_j^0| \leq \theta\}$ and $\mathcal{A}_\infty = \{P \in \mathbb{R}_+^J \mid d_\infty(P, P_0) = \|P - P_0\|_\infty \leq \theta\} = \{P \in \mathbb{R}_+^J \mid \max_{1 \leq j \leq J} |p_j - p_j^0| \leq \theta\}$.

Next, for these two sets, the value θ is determined by the amount of given historical data and the confidence level, i.e., how much percentage is required for the distance between the true distribution and the empirical distribution to be less than θ . For example, if the confidence level is equal to 95%, then the constraint describing \mathcal{A}_1 (resp. \mathcal{A}_∞) guarantees that the true distribution is within \mathcal{A}_1 (resp. \mathcal{A}_∞) with at least 95% confidence level. Here θ is called the distance level, which can be calculated based on the given confidence level and the amount of historical data. Intuitively, the more the historical data is available, the smaller the distance between the empirical distribution and the true distribution. Therefore, for a confidence set with a fixed confidence level, the more the historical data is available, the smaller the value of θ . It follows that this confidence set shrinks. Meanwhile, similar to the related studies as described in [33] and [34], we have the following convergence rates available:

- 1) Corresponding to the given set of historical data (with S samples) and J bins, the convergence rate between P and P_0 under L_1 norm is described as follows:

$$\Pr\{\|P - P_0\|_1 \leq \theta\} \geq 1 - 2Je^{-\frac{2S\theta}{J}}. \quad (18)$$

- 2) Corresponding to the given set of historical data (with S samples) and J bins, the convergence rate between P and P_0 under L_∞ norm is described as follows:

$$\Pr\{\|P - P_0\|_\infty \leq \theta\} \geq 1 - 2Je^{-2S\theta}. \quad (19)$$

Based on the above descriptions, assuming the confidence level to be γ , i.e., the right-hand sides of inequalities (18) and (19) are γ , i.e., $\gamma = 1 - 2Je^{-\frac{2S\theta}{J}}$ under L_1 norm and $\gamma = 1 - 2Je^{-2S\theta}$ under L_∞ norm. It follows that the values of θ with respect to different norms can be obtained as follows:

$$\theta \text{ for } L_1 \text{ norm:} \quad \theta_1 = \frac{J}{2S} \log \frac{2J}{1-\gamma}, \quad (20)$$

$$\theta \text{ for } L_\infty \text{ norm:} \quad \theta_\infty = \frac{1}{2S} \log \frac{2J}{1-\gamma}. \quad (21)$$

From (20) and (21), it is easy to observe that as the size of historical data S increases to infinity, both θ_1 and θ_∞ decrease to 0. It follows that the confidence sets \mathcal{A}_1 and \mathcal{A}_∞ become singletons and that the corresponding two-stage risk-averse stochastic self-scheduling problem converges to

the traditional two-stage risk-neutral stochastic self-scheduling problem. Note here that the historical data in this paper are used to construct the confidence set, which is incorporated in the decision making/optimization model (15), rather than forecasting the value or distribution of the uncertain parameter. For detailed forecasting methods, the readers are referred to [35], [36], [37], and [38], among others.

V. BENDER'S DECOMPOSITION ALGORITHM

To solve the problem presented in (15) - (17), the Bender's decomposition framework [39] is applied. Due to the independence of scenarios $\xi^1, \xi^2, \dots, \xi^J$, for equation (17), the outer summation and inner maximization part inherited from (16) can be interchanged. Thus, the problem can be reformulated as follows.

$$\begin{aligned} \max_{y, u, v} & - \sum_{t=1}^T \sum_{i \in \mathcal{G}^{\text{CT}}} (u_i^t + v_i^t) \\ & + \min_{P \in \mathcal{A}} \max_{x, \phi} \sum_{j=1}^J p_j \left(\sum_{t=1}^T \sum_{m \in \mathcal{M}} (q_t(\xi^j) x_m^t(\xi^j) - \phi_m^t(\xi^j)) \right) \\ \text{s.t.} & (1) - (8), (9) - (13), \forall \xi; \\ & y_e^t \in \{0, 1\}, \forall e \in \mathcal{H}, \forall t; \\ & \phi_m^t(\xi) \text{ free}, x_m^t(\xi) \geq 0, \forall m \in \mathcal{M}, \forall t. \end{aligned} \quad (22)$$

A. Bender's Decomposition Framework

Before describing the solution approach in detail to solve the problem above, in this subsection the fundamental concepts and steps of Bender's decomposition framework are firstly introduced. Bender's decomposition is a solution algorithm to solve the large-scale (mixed-integer) linear program by partitioning the original problem into a small master problem and a subproblem, both of which are solved in an iterative process. For each iteration, new constraints are generated after solving the sub-problem and then are added to the master problem, which is solved again towards the final optimal solution. For simplicity, a mixed-integer linear program is considered as follows.

$$\begin{aligned} \text{(MIP)} \quad \min_{\varpi, \zeta} & c^\top \varpi + f^\top \zeta \\ \text{s.t.} & A\varpi + B\zeta \geq b, \\ & \zeta \in \Delta, \varpi \geq 0, \end{aligned}$$

where ϖ is a continuous variable, ζ is an integer variable, and Δ represents an integer set.

If ζ is fixed to a feasible integer solution (denoted as $\bar{\zeta}$), the resulting model to solve is the following left model and its corresponding dual form is described in the right hand side:

$$\begin{aligned} \text{(SP)} \quad \min_{\varpi} & c^\top \varpi & \text{(SP-D)} \quad \max_{\tau} & (b - B\bar{\zeta})^\top \tau \\ \text{s.t.} & A\varpi \geq b - B\bar{\zeta}, & \text{s.t.} & A^\top \tau \leq c, \\ & \varpi \geq 0, & & \tau \geq 0, \end{aligned}$$

Therefore, the master problem can be set as $\max_{\zeta} \{f^\top \zeta + \vartheta | \zeta \in \Delta, \text{ new constraints/cuts, } \vartheta \text{ free}\}$, where new constraints/cuts are added after each iteration in which the subproblem (SP-D) is solved and certain conditions are satisfied. The detailed algorithmic steps can be described as follows.

- Initialization: Let $\bar{\zeta} :=$ initial feasible solution, the problem's lower bound $\text{LB} := -\infty$, and the problem's upper bound $\text{UB} := \infty$.
- Step 1: Solve the subproblem (SP-D). If (SP-D) is unbounded, then get unbounded ray $\bar{\tau}$ and add feasibility cut $(b - B\bar{\zeta})^\top \bar{\tau} \leq 0$ to the master problem; otherwise, get optimal solution $\bar{\tau}$, add optimality cut $\vartheta \geq (b - B\bar{\zeta})^\top \bar{\tau}$ to the master problem, and set $\text{UB} := \min\{\text{UB}, f^\top \bar{\zeta} + (b - B\bar{\zeta})^\top \bar{\tau}\}$.
- Step 2: Solve the master problem and get the optimal solutions $\bar{\zeta}$ and $\bar{\vartheta}$. Set $\text{LB} := \max\{\text{LB}, f^\top \bar{\zeta} + \bar{\vartheta}\}$.
- If $\text{UB} - \text{LB} < \epsilon$, the current solution is optimal and stop; otherwise go to Step 1.

In the following subsections, through following this Bender's decomposition framework, a solution approach is derived to solve the original problem (22) by first describing the master problem in Subsection V-B, then detailing the subproblem in Subsection V-C, and finally providing the reformulation techniques and the steps on how to add the optimality cuts in Subsection V-D.

B. Master Problem

Through the decomposition framework, denote z as the second-stage objective and build the master problem in the following (23). In each iteration, the master problem is firstly solved and provides the solutions to the optimality check subproblem, where the corresponding optimality cut would be typically obtained and added to the master problem. Note here that for any solution provided in the master problem, there is no feasibility issue in the second-stage subproblem since a feasible solution based on the given first-stage unit commitment decision can always be found. In other words, there is no need to add feasibility cuts to the master problem. The master problem can be summarized as follows:

$$\begin{aligned} \max_{\substack{y \in \{0,1\} \\ z, u, v}} & - \sum_{t=1}^T \sum_{i \in \mathcal{G}^{\text{CT}}} (u_i^t + v_i^t) + z & (23) \\ \text{s.t.} & (1) - (8), \\ & \text{Optimality cuts.} \end{aligned}$$

C. Optimality Check Subproblem

From master problem (23), given the first-stage solution (y, z) , the second-stage subproblem can be described as:

$$\min_{P \in \mathcal{A}} \max_{x, \phi} \sum_{j=1}^J p_j \left(\sum_{t=1}^T \sum_{m \in \mathcal{M}} (q_t(\xi^j) x_m^t(\xi^j) - \phi_m^t(\xi^j)) \right) \quad (24)$$

$$\text{s.t.} \quad -\beta_m^n x_m^t(\xi^j) + \phi_m^t(\xi^j) \geq \alpha_m^n \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right), \forall j, t, m, \quad (25)$$

$$x_m^t(\xi^j) \geq \underline{c}_m \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right), \forall j, t, m, \quad (26)$$

$$-x_m^t(\xi^j) \geq -\bar{c}_m \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right), \forall j, t, m, \quad (27)$$

$$\begin{aligned} - \sum_{m \in \mathcal{M}} x_m^t(\xi^j) + \sum_{m \in \mathcal{M}} x_m^{t-1}(\xi^j) \\ \geq -V_e^+ y_e^t - \hat{C}(1 - y_e^t), \forall j, t, e, \end{aligned} \quad (28)$$

$$\begin{aligned} - \sum_{m \in \mathcal{M}} x_m^{t-1}(\xi^j) + \sum_{m \in \mathcal{M}} x_m^t(\xi^j) \\ \geq -V_e^- y_e^t - \hat{C}(1 - y_e^t), \forall j, t, e, \end{aligned} \quad (29)$$

$$\phi_m^t(\xi) \text{ free, } x_m^t(\xi) \geq 0, \forall j, t, m. \quad (30)$$

To solve this min-max problem, the inner maximization problem is dualized, with dual variables $\delta_{jtmn}, \lambda_{jtm}^+, \lambda_{jtm}^-, \eta_{jte}^+$, and η_{jte}^- corresponding to constraints (25), (26), (27), (28), and (29), respectively. Then the two minimization parts are combined together as follows (subproblem):

$$\begin{aligned} \psi(y) = \min_{P, \delta, \lambda, \eta} \sum_{j=1}^J \sum_{t=1}^T \left\{ \sum_{m \in \mathcal{M}} \sum_{n=1}^N \delta_{jtmn} \alpha_m^n \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right) \right. \\ + \sum_{m \in \mathcal{M}} \left(\lambda_{jtm}^+ \underline{C}_m \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right) \right. \\ \left. \left. - \lambda_{jtm}^- \overline{C}_m \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right) \right) \right. \\ \left. + \sum_{e \in \mathcal{H}} \left(\eta_{jte}^+ (-V_e^+ y_e^t - \hat{C}(1 - y_e^t)) \right) \right. \\ \left. + \eta_{jte}^- (-V_e^- y_e^t - \hat{C}(1 - y_e^t)) \right\} \\ \text{s.t. } \sum_{j=1}^J p_j = 1, \quad P \in \mathcal{A}, \quad \sum_{n=1}^N \delta_{jtmn} = -p_j, \forall j, t, m, \\ \sum_{n=1}^N \delta_{jtmn} (-\beta_m^n) + \lambda_{jtm}^+ - \lambda_{jtm}^- + \sum_{e \in \mathcal{H}} (-\eta_{jte}^+ \\ + \eta_{j(t+1)e}^+ + \eta_{jte}^- - \eta_{j(t+1)e}^-) \geq p_j q_j^t, \forall j, t, m, \\ \delta_{jtmn}, \lambda_{jtm}^+, \lambda_{jtm}^-, \eta_{jte}^+, \eta_{jte}^- \leq 0, \forall j, t, m, n. \end{aligned}$$

D. Reformulation Techniques

To characterize the constraints $P \in \mathcal{A}$, the reformulation techniques in the following are introduced so that the subproblem above can be transformed into a mixed-integer linear programming problem.

1) L_1 Norm Case: For the L_1 norm case, $P \in \mathcal{A}$ represents $\sum_{j=1}^J |p_j - p_j^0| \leq \theta$, which is equivalent to

$$\begin{aligned} \sum_{j=1}^J w_j &\leq \theta, \\ w_j &\geq p_j - p_j^0, \forall j = 1, \dots, J, \\ w_j &\geq p_j^0 - p_j, \forall j = 1, \dots, J. \end{aligned}$$

2) L_∞ Norm Case: For the L_∞ norm case, $P \in \mathcal{A}$ represents $\max_{1 \leq j \leq J} |p_j - p_j^0| \leq \theta$, which is equivalent to

$$|p_j - p_j^0| \leq \theta, \forall j = 1, \dots, J.$$

Therefore, the second-stage subproblem can be reformulated as an MILP problem, with the first-stage solution (y, z) given. After solving the problem above, the solutions δ, λ , and η and the value $\psi(y)$ are obtained.

- 1) If $\psi(y) \geq z$, the master problem is optimal;
- 2) If $\psi(y) < z$, generate a corresponding optimality cut in the following form and add it to the master problem,

$$\begin{aligned} \sum_{j=1}^J \sum_{t=1}^T \left\{ \sum_{m \in \mathcal{M}} \sum_{n=1}^N \delta_{jtmn} \alpha_m^n \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right) \right. \\ + \sum_{m \in \mathcal{M}} \left(\lambda_{jtm}^+ \underline{C}_m \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right) \right. \\ \left. \left. - \lambda_{jtm}^- \overline{C}_m \left(\sum_{e \in (\mathcal{H}_m^{\text{in}} \cup \mathcal{H}_m^{\text{sl}})} y_e^t \right) \right) \right. \\ \left. + \sum_{e \in \mathcal{H}} \left(\eta_{jte}^+ (-V_e^+ y_e^t - \hat{C}(1 - y_e^t)) + \eta_{jte}^- (-V_e^- y_e^t - \hat{C}(1 - y_e^t)) \right) \right\} \geq z. \end{aligned}$$

VI. COMPUTATIONAL RESULTS

In this section, the proposed approach is implemented for a CCU in the IEEE 118-bus system available online at <http://motor.ece.iit.edu/data>. This CCU consists of two combustion turbines and one steam turbine (2CTs + 1ST). In addition, the time horizon is set to be 24 hours. All the experiments were carried out using a computer with Intel Dual Core 2.60 GHz and 8 GB memory. CPLEX 12.5 via C++ Concert Technology was applied to solve the problem under its default setting.

A. Data Setting

For the CCU, its physical parameters are presented as shown in Table II, where the first column shows the mode indices and for each mode, the remaining columns show the combination of CTs and STs of this mode, its generation lower bound, generation upper bound, ramp-up rate limit, ramp-down rate limit, and generation cost function coefficients of the quadratic, linear, and constant terms, respectively.

TABLE II
PHYSICAL PARAMETERS

Mode No	mCTs +nST	\underline{C} (MW)	\overline{C} (MW)	V^+ (MW/h)	V^- (MW/h)	a (\$/MW ² h)	b (\$/MWh)	c (\$/h)
0	0+0	0	0	0	0	0	0	0
1,2	1+0	5	25	25	25	0.025	34.39	16.54
3,4	1+1	10	37.5	37.5	37.5	0.014	61.30	14.33
5	2+0	10	50	50	50	0.013	68.78	16.54
6	2+1	15	75	75	75	0.010	134.13	13.51

For the price data, the number of bins is set as $J = 5$ and each individual data is generated through Monte Carlo sampling based on the forecasted price (see Fig. 2), which is generated based on the PJM real-time price in 2015 [40].

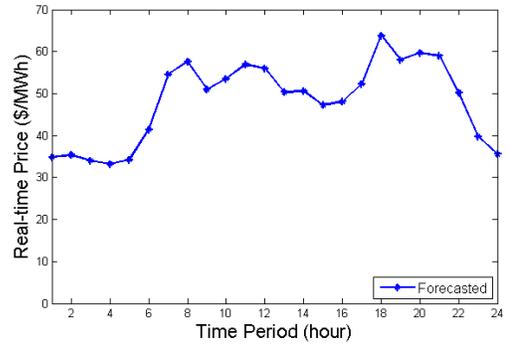


Fig. 2. Price Evolution Over Time

B. Effects of Historical Data

This experiment first tests the effects of historical data on the conservatism of the proposed model, i.e., how the objective value and θ change as the amount of available historical data varies. Let the confidence level γ fixed at 99% and allow the amount of historical data to vary in the range [5, 10000]. In addition, the results are obtained by comparing three models: (1) data-driven risk-averse stochastic self-scheduling with L_1 norm (denoted as DD-1), (2) data-driven risk-averse stochastic

self-scheduling with L_∞ norm (denoted as DD-Inf), and (3) risk-neutral stochastic self-scheduling, i.e., the traditional two-stage stochastic self-scheduling (denoted as SS), as shown in Table III and Figures 3 - 4. Meanwhile, Table III reports the time (in seconds) to solve each model, as shown in the column labeled “Time(s)”.

TABLE III
EFFECTS OF HISTORICAL DATA

# of data	DD-1			DD-Inf			SS	
	OBJ(\$)	θ	Time(s)	OBJ(\$)	θ	Time(s)	OBJ	Time(s)
5	31711.1	3.4539	12.0	31868.6	0.6908	12.2	57235.65	3.7
50	39598.6	0.3454	8.3	49090.5	0.0691	8.1	57235.65	3.7
100	48417.1	0.1727	8.6	53169	0.0345	8.9	57235.65	3.7
500	55474	0.0345	8.7	56422.3	0.0069	8.9	57235.65	3.7
1000	56352.3	0.0173	8.9	56823.1	0.0035	8.7	57235.65	3.7
2000	56796.5	0.0086	8.7	57035.3	0.0017	8.8	57235.65	3.7
5000	57056.9	0.0035	8.5	57153.1	0.0007	8.5	57235.65	3.7
10000	57148.8	0.0017	8.5	57200.3	0.0003	8.7	57235.65	3.7

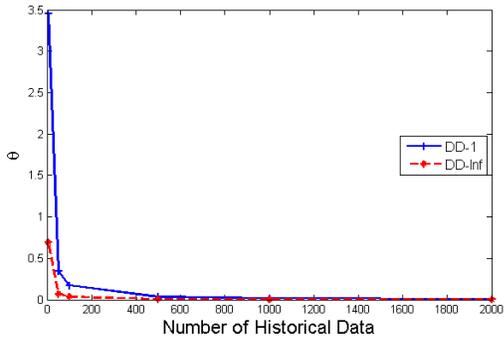


Fig. 3. Effects of Historical Data on the Value of θ

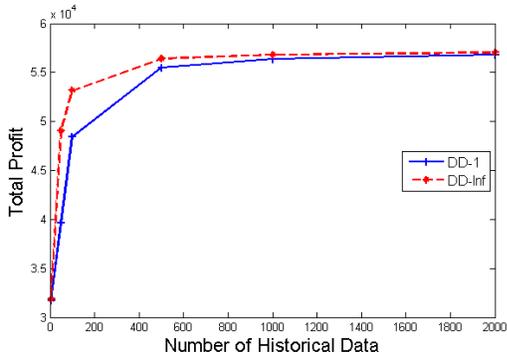


Fig. 4. Effects of Historical Data on the Total Profit

From Table III and Figures 3 - 4, it can be observed that, as the amount of historical data (i.e., S) increases, the value of θ decreases and the total profit increases with respect to both norms used in this paper. The reason is that the confidence set of the true distribution shrinks as the amount of historical data increases, and it follows that the value of θ decreases. Also, the problem becomes less conservative and thus the total profit increases and converges to the objective value of the risk-neutral stochastic optimization model. From Figures 3 - 4, it can be observed that this convergence is gained very fast

(e.g., almost converge when $S = 500$) for the proposed model. Furthermore, given the same amount of data, the model DD-1 is more conservative than DD-Inf, since L_1 norm leads to a larger confidence set than L_∞ norm.

In addition, to further investigate how extra data can help reduce the conservatism of the proposed model corresponding to a given amount of historical data, referred as the value of data, the results are reported in Table IV. Given s historical data, two gaps (labeled “Gap” in Table IV) are defined as follows:

$$\text{Gap}_1(s) = \text{OBJ}^0 - \text{OBJ}_1(s), \quad (31)$$

$$\text{Gap}_\infty(s) = \text{OBJ}^0 - \text{OBJ}_\infty(s), \quad (32)$$

where OBJ^0 , $\text{OBJ}_1(s)$, and $\text{OBJ}_\infty(s)$ are the objective values obtained from the models SS, DD-1, and DD-Inf, respectively. Based on these gaps, the value of data (labeled “VoD” in Table IV) are defined as follows:

$$\text{VoD}_1(s, \bar{s}) = \frac{\text{Gap}_1(s) - \text{Gap}_1(\bar{s})}{\bar{s} - s}, \quad \text{for } \bar{s} > s, \quad (33)$$

$$\text{VoD}_\infty(s, \bar{s}) = \frac{\text{Gap}_\infty(s) - \text{Gap}_\infty(\bar{s})}{\bar{s} - s}, \quad \text{for } \bar{s} > s. \quad (34)$$

The value of data indicates the decrement of the gap between the risk-averse stochastic optimization objective value and the risk-neutral stochastic optimization objective value by collecting additional data. For instance, in Table IV, $\text{VoD}_1(5, 50) = (\text{Gap}_1(5) - \text{Gap}_1(50))/(50 - 5) = (25524.5 - 17637)/45 = 175.28$ and $\text{VoD}_\infty(5000, 10000) = (\text{Gap}_\infty(5000) - \text{Gap}_\infty(10000))/(10000 - 5000) = (82.5 - 35.3)/5000 = 0.01$.

From Table IV, it is easy to observe that both “Gap” and “VoD” decrease as the amount of historical data increases. The value of data is even less than 1 after there are already 1000 historical data, for both DD-1 and DD-Inf models. That means the proposed risk-averse model converges to the risk-neutral model so quickly that not too much data are needed to obtain a solution close to that of the risk-neutral one with the same confidence guarantee.

TABLE IV
VALUE OF DATA

# of data	DD-1		DD-Inf	
	Gap	VoD	Gap	VoD
5	25524.5		25367	
50	17637	175.28	8145.1	382.71
100	8818.5	176.37	4066.6	81.57
500	1761.6	17.64	813.3	8.13
1000	883.3	1.76	412.5	0.80
2000	439.1	0.44	200.3	0.21
5000	178.7	0.09	82.5	0.04
10000	86.8	0.02	35.3	0.01

C. Effects of Confidence Level

This experiment tests the effects of confidence level γ on the conservatism of the proposed model, as the change of γ leads to the change of θ from (20) and (21). The amount of historical data (i.e., S) is set as 500 and γ can be varied in the range $[0.5, 0.99]$. Then this experiment shows how the

objective values and θ change as γ changes for both DD-1 and DD-Inf models in Table V. Meanwhile, the time (in seconds) to solve each model is reported, as shown in the column labeled “Time(s)”.

TABLE V
EFFECTS OF CONFIDENCE SET

γ	DD-1			DD-Inf		
	OBJ(\$)	θ	Time(s)	OBJ(\$)	θ	Time(s)
0.5	56469.7	0.0150	8.9	56882	0.0030	9.2
0.6	56413.5	0.0161	8.9	56858.4	0.0032	8.6
0.7	56342.1	0.0175	8.8	56823.1	0.0035	8.7
0.8	56234.8	0.0196	8.5	56775.9	0.0039	8.3
0.9	56061.2	0.0230	8.8	56693.4	0.0046	8.9
0.95	55882.5	0.0265	8.6	56610.9	0.0053	9.5
0.99	55474	0.0345	8.7	56422.3	0.0069	8.9

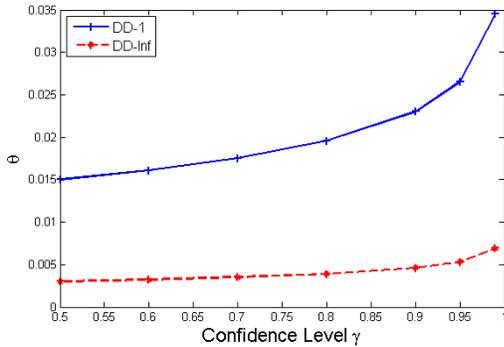


Fig. 5. Effects of Confidence Level on the Value of θ

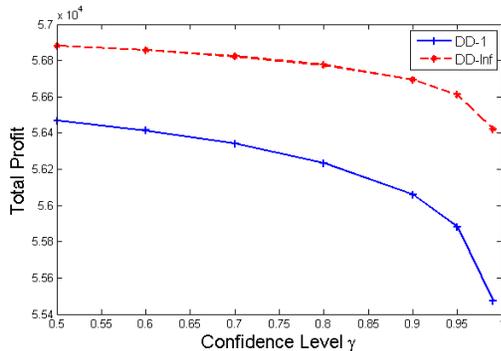


Fig. 6. Effects of Confidence Level on the Total Profit

From Table V and Figures 5 - 6, it can be observed that, as the confidence level increases, the value of θ increases and the objective value decreases for both models. The reason is that with the confidence level increasing, a larger confidence set is needed to ensure that the true distribution is within the confidence set with a higher confidence level (chance). As a result, the model becomes more conservative and thus the objective value (i.e. total profit) decreases. Similar to the results in Table III, DD-1 model is more conservative than DD-Inf model, as L_1 norm leads to a smaller convergence rate than L_∞ norm.

D. Comparisons with Both Stochastic and Robust Optimization Approaches

In this subsection, with further tests on more variations, the proposed approach is compared with traditional two-stage stochastic optimization approach and traditional two-stage robust optimization approach (denoted as RS) together. The confidence level γ is fixed at 95% and let the number of data increases from 5 to 500. For fair comparison, the online/offline status of each edge in the transition graph (which is the first-stage decision of each model) is first obtained from each model, then through fixing the edge status at each time period, the second-stage problem is solved for each model under different distributions in two cases, i.e., the worst-case distribution (which is obtained from the data-driven model) and a randomly generated distribution from the confidence set, respectively. The results are reported in Tables VI and VII. In both tables, the column labeled “Trans(\$)” represents the transition cost of the CCU obtained from the first-stage problem, the columns labeled “Worst(\$)” and “Rand(\$)” represent the total profits under the worst-case distribution and the randomly generated distribution, respectively, and the column labeled “Time(s)” represents the time to solve each model.

From Tables VI and VII, it can be observed that under each setting of θ , the proposed data-driven approach provides a higher profit than both SS and RS under both the worst-case distribution and randomly generated distribution. In addition, the data-driven model has a little higher transition cost than SS and lower than RS. The reason is that data-driven model results in more transitions than SS does to accommodate the electricity price uncertainty, while RS is very conservative and leads to a even higher transition cost. It indicates that the data-driven model provides more reliable solutions than SS and less conservative solutions than RS.

VII. CONCLUSIONS

In this paper, a two-stage data-driven risk-averse stochastic self-scheduling model is proposed for combined-cycle units participating in the real-time market under price uncertainty. The model utilizes the historical data and does not require the precise price distribution as the traditional two-stage stochastic model does. Instead, the confidence set of the unknown price distribution is constructed based on the empirical distribution obtained from the historical price data. It can be shown that corresponding to a given confidence level, as the amount of historical data increases to infinity, the conservatism of the proposed model vanishes and converges from risk-averse to risk-neutral. Thus, the proposed model is data-driven and its conservatism can be adjusted depending on the amount of data and the value of confidence level. Furthermore, with a given amount of historical data, the value of data is also explored to show how additional data can help reduce the conservatism of the proposed model, while maintaining the same confidence level guarantee.

REFERENCES

- [1] A. L. Ott, “Experience with PJM market operation, system design, and implementation,” *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 528–534, 2003.

TABLE VI
DD-1 vs SS vs RS

# of data	θ	DD-1				SS				RS			
		Trans(\$)	Worst(\$)	Rand(\$)	Time(s)	Trans(\$)	Worst(\$)	Rand(\$)	Time(s)	Trans(\$)	Worst(\$)	Rand(\$)	Time(s)
5	2.6492	590	31711.1	32599.0	11.7	540	31552.5	32467.6	3.7	690	30708.5	32359.7	10.6
10	1.3246	590	31909.7	32711.9	12.1	540	31870.8	32680.2	3.6	690	30708.5	32588.7	10.7
25	0.5298	580	35694.9	36140.6	8.5	540	35673.1	36121.0	3.7	690	30708.5	34607.6	10.4
50	0.2649	580	43709.2	44108.8	8.4	540	43694.4	44094.0	3.5	690	30708.5	41817.1	10.4
100	0.1325	580	50469.8	50883.3	8.5	540	50468.0	50872.7	3.6	690	30708.5	48175.4	10.8
200	0.0662	550	53855.3	54243.1	8.6	540	53853.1	54235.1	3.7	690	30708.5	53472.3	10.5
500	0.0265	550	55882.5	56134.6	8.8	540	55880.2	56133.7	3.7	690	30708.5	55086.5	10.3

TABLE VII
DD-INF vs SS vs RS

# of data	θ	DD-Inf				SS				RS			
		Trans(\$)	Worst(\$)	Rand(\$)	Time(s)	Trans(\$)	Worst(\$)	Rand(\$)	Time(s)	Trans(\$)	Worst(\$)	Rand(\$)	Time(s)
5	0.5298	580	32100.8	33337.5	8.4	540	32045.7	33211.8	3.6	690	30708.5	32554.1	10.3
10	0.2649	580	32863.4	33797.7	8.3	540	32828.7	33741.4	3.5	690	30708.5	33302.9	10.6
25	0.1060	580	44740.9	46191.0	8.5	540	44719.2	46165.2	3.7	690	30708.5	41051.7	10.5
50	0.0530	550	50988.3	52350.2	8.9	540	50970.5	52325.9	3.6	690	30708.5	48352.7	10.7
100	0.0265	550	54112	55030.0	8.6	540	54109.8	55013.4	3.6	690	30708.5	52029.5	10.6
200	0.0132	550	55679.7	56308.0	8.6	540	55678.2	56303.1	3.7	690	30708.5	54921.9	10.3
500	0.0053	540	56610.9	56613.3	8.8	540	56610.9	56612.7	3.8	690	30708.5	55850.3	10.8

- [2] "Energy and Operating Reserve Markets," <https://www.misoenergy.org/LIBRARY/BUSINESSPRACTICESMANUALS/Pages/BusinessPracticesManuals.aspx>, 2015.
- [3] A. Papavasiliou, Y. He, and A. Svoboda, "Self-commitment of combined cycle units under electricity price uncertainty," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1690–1701, 2015.
- [4] T. Li and M. Shahidehpour, "Price-based unit commitment: a case of lagrangian relaxation versus mixed integer programming," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 2015–2025, 2005.
- [5] T. Li, M. Shahidehpour, and Z. Li, "Risk-constrained bidding strategy with stochastic unit commitment," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 449–458, 2007.
- [6] J. M. Arroyo and A. J. Conejo, "Optimal response of a thermal unit to an electricity spot market," *IEEE Trans. Power Syst.*, vol. 15, no. 3, pp. 1098–1104, 2000.
- [7] A. J. Conejo, J. M. Arroyo, J. Contreras, and F. A. Villamor, "Self-scheduling of a hydro producer in a pool-based electricity market," *IEEE Trans. Power Syst.*, vol. 17, no. 4, pp. 1265–1272, 2002.
- [8] A. J. Conejo, F. J. Nogales, J. M. Arroyo, and R. Garcia-Bertrand, "Risk-constrained self-scheduling of a thermal power producer," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1569–1574, 2004.
- [9] H. Y. Yamin and S. M. Shahidehpour, "Self-scheduling and energy bidding in competitive electricity markets," *Electric Power Syst. Res.*, vol. 71, no. 3, pp. 203–209, 2004.
- [10] M. Tamayo, X. Yu, X. Wang, and J. Zhang, "Configuration based combined cycle model in market resource commitment," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2013, pp. 1–5.
- [11] A. L. Ott, "Evolution of computing requirements in the pjm market: Past and future," in *Proc. IEEE Power Energy Soc. Gen. Meeting*. IEEE, 2010, pp. 1–4.
- [12] H. Hui, C.-N. Yu, F. Gao, and R. Surendran, "Combined cycle resource scheduling in ercot nodal market," in *Proc. IEEE Power Energy Soc. Gen. Meeting*. IEEE, 2011, pp. 1–8.
- [13] B. Lu and M. Shahidehpour, "Short-term scheduling of combined cycle units," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1616–1625, 2004.
- [14] C. Liu, M. Shahidehpour, Z. Li, and M. Fotuhi-Firuzabad, "Component and mode models for the short-term scheduling of combined-cycle units," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 976–990, 2009.
- [15] E. Ni, P. B. Luh, and S. Rourke, "Optimal integrated generation bidding and scheduling with risk management under a deregulated power market," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 600–609, 2004.
- [16] M. A. Plazas, A. J. Conejo, and F. J. Prieto, "Multimarket optimal bidding for a power producer," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 2041–2050, 2005.
- [17] A. Street, F. Oliveira, and J. M. Arroyo, "Contingency-constrained unit commitment with security criterion: A robust optimization approach," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1581–1590, 2011.
- [18] R. Jiang, J. Wang, and Y. Guan, "Robust unit commitment with wind power and pumped storage hydro," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 800–810, 2012.
- [19] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 52–63, 2013.
- [20] L. Baringo and A. J. Conejo, "Offering strategy via robust optimization," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1418–1425, 2011.
- [21] M. Asensio and J. Contreras, "Stochastic unit commitment in isolated systems with renewable penetration under cvar assessment," *IEEE Trans. on Smart Grid*, vol. 7, no. 3, pp. 1356–1367, 2016.
- [22] I. Momber, A. Siddiqui, T. G. San Román, and L. Söder, "Risk averse scheduling by a pev aggregator under uncertainty," *IEEE Trans. on Power Syst.*, vol. 30, no. 2, pp. 882–891, 2015.
- [23] F. Farzan, M. A. Jafari, R. Masiello, and Y. Lu, "Toward optimal day-ahead scheduling and operation control of microgrids under uncertainty," *IEEE Trans. on Smart Grid*, vol. 6, no. 2, pp. 499–507, 2015.
- [24] P. Jorion, *Value at risk*. McGraw-Hill, New York, 1997.
- [25] M. Shafie-khah, E. Heydarian-Forushani, M. E. H. Golshan, M. P. Moghaddam, M. K. Sheikh-El-Eslami, and J. P. Catalão, "Strategic offering for a price-maker wind power producer in oligopoly markets considering demand response exchange," *IEEE Trans. on Ind. Informat.*, vol. 11, no. 6, pp. 1542–1553, 2015.
- [26] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, pp. 21–42, 2000.
- [27] M. Grbovic, W. Li, N. A. Subrahmanya, A. K. Usadi, and S. Vucetic, "Cold start approach for data-driven fault detection," *IEEE Trans. on Ind. Informat.*, vol. 9, no. 4, pp. 2264–2273, 2013.
- [28] H. A. Khorshidi, I. Gunawan, and M. Y. Ibrahim, "Data-driven system reliability and failure behavior modeling using FMECA," *IEEE Trans. on Ind. Informat.*, vol. 12, no. 3, pp. 1253–1260, 2016.
- [29] D. Wang, J. Liu, and R. Srinivasan, "Data-driven soft sensor approach for quality prediction in a refining process," *IEEE Trans. on Ind. Informat.*, vol. 6, no. 1, pp. 11–17, 2010.
- [30] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K.-L. Tsui, "A two-stage data-driven-based prognostic approach for bearing degradation problem," *IEEE Trans. on Ind. Informat.*, vol. 12, no. 3, pp. 924–932, 2016.
- [31] L. Fan and Y. Guan, "An edge-based formulation for combined-cycle units," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 1809–1819, 2016.
- [32] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*. John Wiley & Sons, 2012.
- [33] C. Zhao and Y. Guan, "Data-driven stochastic unit commitment for integrating wind generation," *IEEE Trans. on Power Syst.*, vol. 31, no. 4, pp. 2587–2596, 2016.
- [34] R. Jiang, Y. Guan, and J.-P. Watson, "Risk-averse stochastic unit commitment with incomplete information," *IIE Transactions*, vol. 48, no. 9, pp. 838–854, 2016.

- [35] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting Methods and Applications*. John Wiley & Sons, 2008.
- [36] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi, "Prediction interval estimation of electricity prices using pso-tuned support vector machines," *IEEE Trans. on Ind. Informat.*, vol. 11, no. 2, pp. 322–331, 2015.
- [37] M. Rafei, T. Niknam, and M.-H. Khooban, "Probabilistic forecasting of hourly electricity price by generalization of elm for usage in improved wavelet neural network," *IEEE Trans. on Ind. Informat.*, vol. 13, no. 1, pp. 71–79, 2017.
- [38] S. Buhan and I. Çadırcı, "Multistage wind-electric power forecast by using a combination of advanced statistical methods," *IEEE Trans. on Ind. Informat.*, vol. 11, no. 5, pp. 1231–1242, 2015.
- [39] J. F. Benders, "Partitioning procedures for solving mixed-variables programming problems," *Numerische Mathematik*, vol. 4, no. 1, pp. 238–252, 1962.
- [40] "Daily Real-Time LMP," <http://www.pjm.com/markets-and-operations/energy/real-time/lmp.aspx>, 2016.



Kai Pan (S'13-M'16) received the B.S. degree in industrial engineering from Zhejiang University, Hangzhou, China, in 2010 and the M.S. and Ph.D. degrees in industrial and systems engineering from the University of Florida, Gainesville, FL, USA, in 2014 and 2016, respectively. He is currently an assistant professor in the Department of

Logistics and Maritime Studies at the Hong Kong Polytechnic University, Hung Hom, Hong Kong. His research interests include stochastic integer programming, data-driven optimization, and energy system applications.



Yongpei Guan (M'10-SM'13) received the dual B.S. degrees in mechanical engineering and economic decision science from Shanghai Jiao Tong University, Shanghai, China, in 1998, the M.Phil. degree in industrial engineering and engineering management from Hong Kong University of Science and

Technology, Hong Kong, China, in 2001, and the Ph.D. degree in industrial and systems engineering from Georgia Institute of Technology, Atlanta, GA, USA, in 2005. He is currently a full professor in the Department of Industrial and Systems Engineering at the University of Florida, Gainesville, FL, USA. His research interests include stochastic and discrete optimization, data-driven stochastic optimization, supply chain management, and renewable energy integration.