Robust Unit Commitment with $n-1$ Security Criteria

Arash Gourtani

School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK

(A.Mostajeran-Gourtani@soton.ac.uk)

Huifu Xu [1]

School of Mathematical Sciences, University of Southampton

SO17 1BJ, Southampton UK

(H.Xu@soton.ac.uk)

David Pozo

Industrial and Systems Engineering Department,

Pontificia Universidad Catlica de Chile Santiago, Chile

(dpozoc@uc.cl)

Tri-Dung Nguyen

School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK

(T.D.Nguyen@soton.ac.uk)

January 15, 2016

Abstract

The short-term unit commitment and reserve scheduling decisions are made in the face of increasing supply-side uncertainty in power systems. This has mainly been caused by a higher penetration of renewable energy generation that is encouraged and enforced by the market and policy makers. In this paper, we propose a two-stage stochastic and distributionally robust modeling framework for the unit commitment problem with supply uncertainty. Based on the availability of the information on the distribution of the random supply, we consider two specific models: (a) a moment model where the mean values of the random supply variables are known, and (b) a mixture distribution model where the true probability distribution lies within the convex hull of a finite set of known distributions. In each case, we reformulate these models through Lagrange dualization as a semi-infinite program in the former case and a one-stage stochastic program in the latter case. We solve the reformulated models using sampling method and sample average approximation, respectively. We also establish exponential rate of convergence of the optimal value when the randomization scheme is applied to discretize the semi-infinite constraints. The proposed robust unit commitment models are applied to an illustrative case study, and numerical test results are reported in comparison with the two-stage non-robust stochastic programming model.

Key words: Unit commitment problem; distributionally robust optimization; mixture distribution; sample average approximation; convergence analysis

---

1 Introduction

The recent increase in the deployment of renewable energy resources such as wind power is having a significant impact on the short-term operational and long-term investment decisions in power systems due to their non-dispatchability and intermittent nature. In the short term, the higher penetration of wind power and the lack of efficient storage facilities have an adverse effect on the stability of generation output. One of the most crucial decision problems that are affected by the short-term supply uncertainty is the unit commitment (UC for short) problem [29, 43]. The objective of the UC problem is to minimize the overall generation cost by determining the hourly unit commitment and the reserve schedule for the day ahead given the demand and wind forecasts.

Classical models for the UC problem are often deterministic and consider supply and demand for electricity in the day ahead to be known in advance. Whilst the demand forecast for the day ahead can be reasonably estimated, the high reliance of the generation output on the unreliable wind power potentially renders optimal solutions of a deterministic model either heavily infeasible or non-optimal under realized supply [27]. Stochastic and robust optimization models provide an alternative approach to incorporating the increased uncertainties associated with the wind and load forecasts into power system operations. In this sense, approaches to account for the uncertainty in renewable energy generation in the UC problem fall into three categories: (two-stage) stochastic programming, chance-constrained stochastic programming, and robust optimization.

The first approach of two-stage stochastic optimization [39] has been used widely for solving the UC problem [30, 44, 46], where energy and reserve generation are jointly scheduled to meet demand under stochastic wind supply. Pozo and Contreras [32] and Qianfan et al. [33] propose a chance-constrained programming approach to deal with the joint energy and reserve scheduling UC where one or several constraints must be satisfied with a given probability. One of the key assumptions in two-stage stochastic programming and chance-constrained programming is that the decision maker has complete information on the distribution of the uncertain parameters. However, limited predictability and high volatility of the renewable supply often make this assumption unrealistic.

In the third approach of classical robust optimization, the only available information on the uncertain parameters is its support, i.e., the set of all scenarios or possible realizations of the unknown parameters [4, 7, 41]. For the basic concept and a thorough survey of robust optimization, we refer the interested readers to review papers by Aissi et al. [1], Kouvelis and Yu [22] and Ben-Tal et al. [5]. In the context of the UC problem, Jiang et al. [20] and Bertsimas et al. [9] provide a robust optimization formulation and an adaptive robust optimization to address the wind power and demand uncertainty, respectively.

The min-max robust optimization approach is often criticized for being over conservative and/or not utilizing available partial information about the distribution of the uncertainty such as the mean and variance. *Distributionally robust optimization*, as part of the robust modeling framework, is considered a powerful remedy where the optimal decision is based on the worst probability distribution rather than worst scenario from a set of distributions constructed through the partial information.

There is extensive research published on UC models where energy and reserve are scheduled together. Table 1 summarizes some of those references that are closely related to the models proposed in this paper.

Table 1: A survey of UC models with security criteria

Reference	Multi-period	Network	Security	Model	Uncertainty	Available information
[10]	no	no	$n-2$	Det. with prob. constr.	contingencies	probability of outage
[3]	no	yes	$n-1$	Det. MILP	contingencies	set of plausible outages
[12]	no	yes	$n-1$	TS-SP* and MILP ref.†	contingencies	probability of outage
[13, 14]	yes	yes	$n-1$	TS-SP and MILP ref.	contingencies	probability of outage
[11]	yes	no	$n-1$	TS-SP and MILP ref.	contingencies	probability of outage
					demand and wind	demand and wind pdf
[42]	yes	no	$n-K$	Det. with WC§ contingency	contingencies	set of plausible outages
[34]	yes	no	$n-1$	TS-SP and MILP ref.	demand and wind	wind and demand pdf
[21]	yes	yes	$n-1$	TS-SP and MILP ref.	contingencies	probability of outage
[9]	yes	yes	$n-1$	Det. ARO‡	demand	demand bound uncertainty
[32]	yes	no	$n-K$	CC-SP♦ and MILP ref.	demand and wind	demand and wind pdf
Current paper	yes	no	$n-1$	distributionally robust	wind	distributional information

∗ → Two-stage stochastic programming † → Mixed integer linear programming reformulation § → Worst-case

‡ → Adaptive robust optimization ♦ → Chance-constrained stochastic programming

For a comparison between the existing models in the literature and the proposed distributionally robust UC model in this paper, we provide some classification criteria as follows.

- *The uncertainty sources and their available information.* Contingencies, demand and wind production are some common sources of uncertainty in the UC problem with security criterion. Contingency events are usually considered as scenarios to include a deterministic or stochastic UC constraint. A bunch of post-contingency power flow operation equations are included in the problem to model how the system remains stable under the loss of one or more unit or line (see [3, 10, 12–14, 21, 34]). Probability of the contingencies may be known (e.g. in [10, 12, 21, 34]) or may not be known (e.g. in [3, 32, 42]). Bouffard and Galiana [11], Restrepo and Galiana [34] and Pozo and Contreras [32] include stochastic demand and wind with full knowledge of probability distributions of the uncertainties. Bertsimas et al. [9] include stochastic demand with partial information, where the uncertainty set is defined as box uncertainty with a budget constraint.

  Here we model the wind production as the only uncertain parameter with partial information on its probability distribution such as moments and scope of the distribution (representation as a mixture of some know distributions).

- *The number of simultaneous contingencies.* The purpose of security criteria is to keep the system stable in case of one ($n-1$ criterion) or more outages ($n-K$ criteria) of a generating unit or line, where reserve planning is justified to compensate for possible outages. The $n-1$ criterion has been extensively applied to the UC problem [3, 11–14, 21, 34]. Street et al. [42] and Pozo and Contreras [32] extend the security criterion up to K simultaneous contingencies. However, more strict criteria increases the complexity of the model and its tractability.

  Here we propose a model where a $n-1$ security criterion is included in the sense that if one unit is lost, the demand is met with the scheduled reserve from the first stage under any wind scenario.

- *Modeling of contingencies.* Outages can be treated as deterministic parameters and the preventive actions are taken pre-contingency and through inclusion of deterministic constraints. These security constraints will ensure sufficient resources for the normal operation of the system in the event of a contingency (see Arroyo and Galiana [3]). On the other hand, outages can be treated as stochastic parameters, in which case the objective function includes expected value of the second-stage recourse costs. Contingencies probabilities should be known for this model, and the set of post-contingencies equations, one for each outage, is incorporated into the model (see Bouffard and Galiana [11], Bouffard et al. [13]). Note that, when other sources of uncertainty exist, post-contingencies equations should be extended for each scenario, which may lead to an intractable problem. Because of that, some authors limit the scenarios of contingencies to an umbrella of credible contingencies [3, 11–14, 21]. Another approach to deal with security criteria is posing an optimization problem to determine the worst contingency/contingencies. The works by Street et al. [42] and Pozo and Contreras [32] propose a worst-case optimization problem embedded into a deterministic and chance-constrained UC model, respectively.

  Here we formulate the security criterion as a deterministic constraint for the worst-case outage. We show that this worst-case outage is always the worst in any wind scenario. In this sense, we do not need to add up post-contingency equations for each plausible outage (all generating units) and each wind scenario. There is no cost term in the objective function to account for the extra cost of the corrective actions taken in the event of a contingency. This is a reasonable approach because contingencies have a very low probability of occurrence.

- *Mathematical model and numerical methodology.* Bouffard and Galiana [10] and Arroyo and Galiana [3] propose a deterministic modeling framework, Bouffard and Galiana [11] and Bouffard et al. [13] develop a two-stage stochastic programming model, Pozo and Contreras [32] consider a chance-constrained problem, and Street et al. [42] Bertsimas et al. [9] present a robust optimization formulation. Most of these models are solved through their deterministic counterparts and further as mixed integer programming (MILP for short).

  Here we propose a two stage stochastic programming model to describe the decision making process and then develop a distributionally minimax robust optimization formulation of the two stage stochastic program to address the risks arising from uncertainty of wind power supply. Monte Carlo sampling methods have been applied to solve the resulting mathematical models. Compared to the existing minimax robust optimization model, the distributionally robust minimax model is less conservative.

Bertsimas et al. [9] propose a two-stage deterministic robust optimization model for UC problems, where the uncertainty set is defined through a deterministic set. The solution of the proposed adaptive robust model provides immunity against all realizations of the uncertain data within the deterministic uncertainty set. However, this robust model does not take into account the distributional information of the random variables. In contrast, our model accounts for the available, or partially available, information on the probability distribution of the uncertain data. Furthermore the first-stage decision variables in [9] consist of the on/off commitment variables and the second stage solves an economic dispatch; therefore, they do

not need to schedule reserve for the second-stage uncertainty deviations (North American's UC outlook). However, in our model, the first-stage decision variables are the set of on/off decisions and scheduled energy and reserve (European's UC outlook). Scheduled reserve is used in the second stage as a corrective action to meet the demand under wind production deviations and/or the outage of one generating unit.

Despite the fact that there is a rich body of literature focussing on two-stage stochastic and robust UC with endogenous reserve scheduling and wind-generation models, a distributionally robust UC approach has not been presented yet. The main contributions of this paper are summarized as follows:

- We propose a distributionally robust approach for UC problems with $n-1$ security criteria to tackle the risks arising from day-ahead wind uncertainty. Two specific models are considered: i) a moment model where the mean values of the random wind power supply are known, and ii) a mixture distribution model where the true probability distribution can be represented as a convex combination of some known distributions.

- For the moment model, we reformulate the distributionally minimax two stage robust problem as a semi-infinite programming problem through Lagrange duality and propose a random discretization approach for solving the latter. Under some moderate conditions, we demonstrate exponential rate of convergence of the randomization scheme as sample size increases. For the mixture distribution model, we reformulate the robust model as a one-stage stochastic program through duality, and develop a solution method based on sample average approximation (SAA) to reformulate the problem as a mixed integer linear programming problem (MILP).

- We have undertaken numerical experiments for the proposed new mathematical model carried out comparative analysis with two-stage stochastic UC model in terms of stability of the optimal solutions against variation of the mean and covariance.

The remainder of this paper are organized as follows. Section 2 sets out a standard two-stage stochastic programming model for the unit commitment problem with uncertainty in wind power generation and a distributionally robust formulation by explicitly considering ambiguity of the true probability distribution of the underlying uncertainty. Sections 3 and 4 detail the robust formulations and reformulations by considering some specific structures of the ambiguity set and develop corresponding numerical schemes for solving the mathematical models. Section 5 presents some case studies and comparative analysis of the new models and numerical schemes. Some conclusions are drawn in Section 6.

2 Stochastic unit commitment problem

The unit commitment and generation scheduling problems involve inherent uncertainties stemming from the short-term volatility of demand and unpredictability of wind power. The recent progress in the field of stochastic programming makes it an attractive approach for modeling the UC problem under uncertainty. Research carried out by Carpentier et al. [16], Takriti et al. [43] and Dentcheva and Römisch [18] were amongst the first which formulate the UC problem as a two-stage stochastic program. In this section, we first introduce a two-stage stochastic unit commitment problem (Sto-UC) by taking into account the wind uncertainty and additional technical information. We then further extend the stochastic

framework to include the uncertainty on the distribution of the uncertainty and develop a two-stage distributionally robust model.

2.1 Two-stage stochastic UC with uncertain net load

Consider a unit commitment problem with a set of conventional generating units, denoted by $\mathcal{I} = \{1, \cdots, I\}$, over a time horizon denoted by $\mathcal{T} = \{1, \cdots, T\}$. The net load at time $t \in \mathcal{T}$ is stochastic and denoted by ξ_t. The stochasticity of the load reflects uncertainty of renewable power supply such as wind power. Mathematically, we denote by $\boldsymbol{\xi} = (\xi_t), t \in \mathcal{T}$ a vector of random variables defined over measurable space $(\Omega, \mathcal{F})$ with sigma-algebra. We use Ξ to denote the support set of $\boldsymbol{\xi}$. In this context, we can assume with out loss of generality that Ξ is a compact set (bounded and closed).

A standard two-stage stochastic programming model for the UC problem includes a first-stage decision on the planning of unit power operation before realization of the uncertainty and a second-stage adjustment (recourse action) after the uncertainty is observed. The system operator aims to minimize the total generation costs which comprising the planned generation cost and the expected 'adjustment' cost. Specifically, the framework for the two-stage UC problem can be described as follows.

The first-stage (*here and now*) decisions are taken prior to realization of uncertain net load $\boldsymbol{\xi}$. These include the on/off decisions denoted by u_{it}, for generator i and time period t, i.e.,

$$u_{it} = \begin{cases} 1, & \text{if generator } i \text{ is turned on in time period } t, \\ 0, & \text{otherwise,} \end{cases}$$

the energy dispatch variable q_{it}, and the up and down scheduled reserves r_{it}^{up}, r_{it}^{dw} for generator i at time t. Each generator $i \in \mathcal{I}$ has a fixed on cost of c_i^f and, if on, it has a unit generation cost of c_i^l. The upper and lower generation capacities of generator i are given by $\bar{q}_i$ and $\underline{q}_i$ while the upper and lower limits for reserve up/down for generator i are given by $\bar{r}_{it}^{up}/\bar{r}_{it}^{dw}$ and $\underline{r}_{it}^{up}/\underline{r}_{it}^{dw}$, respectively. The unitary costs of scheduling reserve up $c_i^{r,up}$ and reserve down $c_i^{r,dw}$ for generator i are given as an input of the problem.

The second stage decisions include deployed actual up and down reserves, denoted by $\hat{r}_{it}^{up}(\xi), \hat{r}_{it}^{dw}(\xi)$. The unitary costs for the actual deployment of the reserve up and down are denoted by $\hat{c}_i^{up}$ and $\hat{c}_i^{dw}$.

To avoid imbalance in the supply and demand for energy in the second stage, we introduce additional auxiliary variables for load shedding and wind spillage. The load-shedding variable is denoted by $S_t(\xi)$ and represents the excess demand which cannot be met by the total generation output at time period t and has to be shed at high penalty cost of c_t^{ls}. On the other hand, the wind-spillage variable, denoted by $W_t(\xi)$, is equal to the excess wind power that cannot be utilized upon the realization of the net load. Wind spillage incurs an unitary opportunity cost of c_t^{ws}.

The resulting mathematical model of the two-stage stochastic UC problem with $n - 1$

security criterion and ramping constraints can be presented as follows,

$$(\text{Sto-UC}) \quad \min_{\mathbf{u},\mathbf{q},\mathbf{r}} \quad \sum_i \sum_t \Big[\underbrace{c_i^f u_{it} + c_i^l q_{it}}_{\text{Generation Cost}} + \underbrace{c_i^{r,up} r_{it}^{up} + c_i^{r,dw} r_{it}^{dw}}_{\text{Reserves Cost}} \Big] + \underbrace{\mathbb{E}_P[g(\mathbf{u},\mathbf{q},\mathbf{r},\boldsymbol{\xi})]}_{\text{Expected variation cost}}$$

$$\text{s.t.} \quad q_{it} + r_{it}^{up} \leq \bar{q}_i u_{it}, \ \forall t, i, \tag{2.1}$$

$$q_{it} - r_{it}^{dw} \geq \underline{q}_i u_{it}, \ \forall t, i, \tag{2.2}$$

$$\underline{r}_i^{up} \leq r_{it}^{up} \leq \bar{r}_i^{up}, \ \forall t, i, \tag{2.3}$$

$$\underline{r}_i^{dw} \leq r_{it}^{dw} \leq \bar{r}_i^{dw}, \ \forall t, i, \tag{2.4}$$

$$q_{it}, r_{it}^{dw}, r_{it}^{up} \in \mathbb{R}^+, \ u_{it} \in \{0,1\}, \ \forall t, i, \tag{2.5}$$

where $\mathbb{E}_P$ denotes the mathematical expectation w.r.t. the distribution of $\boldsymbol{\xi}$ over probability space $(P, \Omega, \mathcal{F})$, and $g(\mathbf{u}, \mathbf{q}, \mathbf{r}, \boldsymbol{\xi})$ is the optimal value of the second-stage problem defined as

$$g(\mathbf{u},\mathbf{q},\mathbf{r},\boldsymbol{\xi}): \quad \min_{\hat{\mathbf{r}},\mathbf{W},\mathbf{S}} \quad \sum_i \sum_t \Big\{ \hat{c}_i^{up} \hat{r}_{it}^{up}(\xi) + \hat{c}_i^{dw} \hat{r}_{it}^{dw}(\xi) \Big\} + \sum_t \Big\{ c_t^{ls} S_t(\xi) + c_t^{ws} W_t(\xi) \Big\}$$

$$\text{s.t.} \quad \sum_i \hat{q}_{it}(\xi) + S_t(\xi) - W_t(\xi) = l_t(\xi), \ \forall t, \tag{2.6}$$

$$0 \leq \hat{r}_{it}^{up}(\xi) \leq r_{it}^{up}, \ \forall t, i, \tag{2.7}$$

$$0 \leq \hat{r}_{it}^{dw}(\xi) \leq r_{it}^{dw}, \ \forall t, i, \tag{2.8}$$

$$Q_t - (q_{it} + r_{it}^{up}) + S_t(\xi) \geq l_t(\underline{\xi}), \ \forall t, i, \tag{2.9}$$

$$\hat{q}_{it}(\xi) = q_{it} + \hat{r}_{it}^{up}(\xi) - \hat{r}_{it}^{dw}(\xi), \qquad \forall t, i, \tag{2.10}$$

$$\hat{q}_{it}(\xi) - \hat{q}_{i(t-1)}(\xi) \leq RU_i u_{i(t-1)} + SU_i(1 - u_{i(t-1)}), \ \forall t, i, \tag{2.11}$$

$$\hat{q}_{i(t-1)}(\xi) - \hat{q}_{it}(\xi) \leq RD_i u_{it} + SD_i(1 - u_{it}), \ \forall t, i. \tag{2.12}$$

In the first stage, constraints (2.1) and (2.2) represent the generation limits including the up and down scheduled reserves. The constraints (2.3) and (2.4) bound the minimum and maximum reserves to be scheduled.

In the second-stage problem, constraint (2.6) represents the energy balance for each hour t and net load wind scenario ξ. Constraints (2.7) and (2.8) ensure that the actual up and down reserves used are within the limits of the nominal reserve scheduled in the first stage. Furthermore, constraint (2.9) represents the $n-1$ reliability requirement which ensures that the demand will be met under the failure of up to one generating unit, where Q_t is the total generation upper limit and $\underline{\xi}_t$ is the worst (lowest) realization of wind output at time t. When the time index t is clear from the context, we use $\underline{\xi}$ as a shorthand notation for $\underline{\xi}_t$. For example, $S_t(\underline{\xi})$ is equivalent to $S_t(\underline{\xi}_t)$ and denotes the load shedding that corresponds to the worst realization of the wind output at time t. Constraint (2.10) provides the formulation for the actual power output of unit i at time t and scenario ξ. Finally, constraints (2.11) and (2.12) represent the ramp constraint with RU_i, RD_i, SU_i, SD_i are the ramp up, ramp down, starting up and starting down ramps, respectively. Details about how the reliability and $n-1$ security constraints and the ramp constraints are formed are presented in Appendix A.

2.2 Distributionally robust UC problem

Under the stochastic programming framework (Sto-UC), we assume that the "true" probability distribution P of the random wind variables is known. In practice, however, such distribution is often unknown and hence it has to be estimated through partial information or subjective judgements. One of the possible ways to deal with this issue is to use available information to construct a set of distributions, denoted by $\mathcal{P}$, in which the true probability distribution is assumed to lie. The robust optimization approach for the two-stage stochastic problem with respect to this ambiguity aims to make a decision which is optimal for the worst probability distribution from $\mathcal{P}$.

To ease the exposition, we write x for the first stage decision variables $(\mathbf{u}, \mathbf{q}, \mathbf{r})$ with feasible domain $\mathcal{X}$. Likewise we write y for the second stage decision variables $(\hat{\mathbf{r}}, \mathbf{W}, \mathbf{S})$. Furthermore, we denote the first-stage cost parameters by $c = (\mathbf{c}^f, \mathbf{c}^l, \mathbf{c}^{r,up}, \mathbf{c}^{r,dw})$ and the second-stage cost parameters by $h = (\hat{\mathbf{c}}^{up}, \hat{\mathbf{c}}^{dw}, \mathbf{c}^{ls}, \mathbf{c}^{ws})$. The corresponding mathematical model can be formulated as

$$
\text{(R-UC)} \qquad \min_{x} \quad c^T x + \sup_{P \in \mathcal{P}} \mathbb{E}_P[g(x, \boldsymbol{\xi})] \tag{2.13}
$$
$$
\text{s.t.} \quad x \in \mathcal{X},
$$

where $g(x, \boldsymbol{\xi})$ is the optimal value of the second-stage problem

$$
g(x, \boldsymbol{\xi}) = \min_{y} \quad h^T y \tag{2.14}
$$
$$
\text{s.t.} \quad y \in \mathcal{Y}(x, \boldsymbol{\xi}),
$$

and $\mathcal{Y}(x, \boldsymbol{\xi})$ is the second-stage feasible set depending on x and $\boldsymbol{\xi}$. In the literature of robust optimization, (2.13) is known as a distributionally robust formulation where the robustness is in the sense that the worst probability distribution rather than the worst scenario of the random vector $\boldsymbol{\xi}$ is taken into account. This kind of robust optimization framework can be traced back to the earlier work by Scarf [37] which was motivated to address incomplete information on the underlying uncertainty in supply chain and inventory control problems. In such problems, historical data may be insufficient to estimate future distribution either because sample size of past demand is too small or because there is a reason to suspect that future demand will come from a different distribution that governing past history. Compared to minimax robust optimization model, the distributionally robust formulation is obviously less conservative and hence more compelling in the circumstances where an optimal decision based on the former model may incur excessive economic and/or computational costs to prevent a rare event. Over the past few decades, DRO models have found many applications in operations research, finance and management sciences, see for instances Bertsimas and Popescu [8], Delage and Ye [17], Goh and Sim [19], Mehrotra and Papp[26], Wiesemann, Kuhn and Rustem [47, 48] and Wozabal [49] for various applications and numerical schemes. In particular, Bertsimas et al [6] propose a distributionally robust formulation of two stage linear programming model with applications in finance and facility location planning.

A key element in (R-UC) is the ambiguity set $\mathcal{P}$. Various statistical methods have been proposed in the literature for constructing ambiguity set. Here we consider two popular ones where $\mathcal{P}$ is constructed through moments and mixture distribution.

3 Moment model and sample approximation approach

In this section, we investigate the robust unit commitment problem where the first moment condition of the underlying random wind generation is known. Let $\boldsymbol{\mu} \in \mathbb{R}^T$ denote the mean of $\boldsymbol{\xi}$. We consider the ambiguity set $\mathcal{P}$ being defined as follows:

$$\mathcal{P} = \{P \in \mathscr{P} : \mathbb{E}_P[\boldsymbol{\xi}] = \boldsymbol{\mu}\}, \tag{3.1}$$

where $\mathscr{P}$ denotes the set of all probability measures of measurable space $(\Xi, \mathscr{B})$ induced by $\boldsymbol{\xi}$. For each fixed $x \in \mathcal{X}$, we consider the worst expected value of $g(x, \boldsymbol{\xi})$ over the ambiguity set $\mathcal{P}$:

$$H(x) := \sup_{P \in \mathcal{P}} \mathbb{E}_P[g(x, \boldsymbol{\xi})]. \tag{3.2}$$

Using the moment conditions, we can write $H(x)$ as the optimal value of the following maximization problem

$$
\begin{aligned}
H(x) = \quad &\sup_{P \in \mathcal{M}^+} \quad \int_\Xi g(x, \boldsymbol{\xi}) P(d\boldsymbol{\xi}), \\
&\text{s.t.} \quad \int_\Xi \xi_t P(d\boldsymbol{\xi}) = \mu_t, \text{ for all } t = 1, \cdots, T, \\
&\qquad \int_\Xi P(d\boldsymbol{\xi}) = 1,
\end{aligned}
\tag{3.3}
$$

where $\mathcal{M}^+$ denotes the set of all non-negative finite measures on measurable space $(\Xi, \mathscr{B})$, and μ_t is the tth component of $\boldsymbol{\mu}$. Problem (3.3) is a typical form of classical moment problem. We refer interested readers to a monograph by Landau [23] for a comprehensive discussion of the historical background of the latter. In order to deal with difficulties associated with solving such an infinite dimensional problem, duality theory is often used; for example, see Rockafellar [35]. Here we follow Shapiro [38, Proposition 3.1] to derive the Lagrange dual associated with the moment problem (3.3).

Proposition 3.1 *For a given $x \in \mathcal{X}$, the Lagrange dual of problem (3.2) is*

$$
\begin{aligned}
H_D(x) := \quad &\min_{\boldsymbol{\alpha}} \quad \alpha_0 + \sum_{t=1}^T \alpha_t \mu_t \\
&\text{s.t.} \quad g(x, \boldsymbol{\xi}) \leq \alpha_0 + \sum_{t=1}^T \alpha_t \xi_t, \text{ for all } \boldsymbol{\xi} \in \Xi,
\end{aligned}
\tag{3.4}
$$

where $\alpha_t \in \mathbb{R}$, $t = 0, 1, \cdots, T$, denotes the dual variables corresponding to moment problem constraints and $\Xi \subset \mathbb{R}^T$ is the support set of $\boldsymbol{\xi}$, and $H_D(x) = H(x)$.

Proof: The derivation of the dual formulation is standard and can be found in Shapiro, Dentcheva and Ruszczyński [39]. Since Ξ is compact and the underlying function are continuous in ξ, the strong duality follows from [39, page 308].

□

Using Proposition 3.1, we can reformulate (R-UC) as

$$
\begin{aligned}
\text{(SIP-UC)} \quad \min_{x,\alpha_0,\boldsymbol{\alpha}} \quad & c^T x + \alpha_0 + \boldsymbol{\alpha}^T \boldsymbol{\mu} \\
\text{s.t.} \quad & x \in \mathcal{X}, \\
& g(x,\boldsymbol{\xi}) \le \alpha_0 + \boldsymbol{\alpha}^T \boldsymbol{\xi}, \text{ for all } \boldsymbol{\xi} \in \Xi.
\end{aligned}
\tag{3.5}
$$

Moreover, through a simple analysis, we can show that problem (3.5) is equivalent to the following program which incorporates the details of the second-stage problem

$$
\begin{aligned}
\min_{x,y(.),\alpha_0,\boldsymbol{\alpha}} \quad & c^T x + \alpha_0 + \boldsymbol{\alpha}^T \boldsymbol{\mu} \\
\text{s.t.} \quad & x \in \mathcal{X}, \\
& h^T y(\boldsymbol{\xi}) \le \alpha_0 + \boldsymbol{\alpha}^T \boldsymbol{\xi}, \text{ for all } \boldsymbol{\xi} \in \Xi, \\
& y(\boldsymbol{\xi}) \in \mathcal{Y}(x,\boldsymbol{\xi}), \text{ for all } \boldsymbol{\xi} \in \Xi.
\end{aligned}
\tag{3.6}
$$

In the rest of the paper, we discuss numerical methods for solving (3.5) or (3.6). If we are able to obtain a closed form of $g(x,\xi)$ and show that g is a linear or a quadratic function of ξ, then we may reformulate the semiinfinite constraints of problem (3.5) as a semidefinite constraint and consequently solve the resulting semidefinite programming (SDP) problem with existing methods for SDP. Indeed, this is the main stream work for distributionally robust optimization, see for instance [47, 48] and references therein. Unfortunately, here g is not linear or quadratic in ξ. Likewise, the underlying functions in the constraints of problem (3.6) are also nonlinear, non-quadratic which makes it impossible to reformulate the semiinfinite constraints as semidefinite constraints. This motivates us to solve (3.6) through other methods

3.1 Sample approximation scheme

One of the well-known solution approaches for semi-infinite programs is random discretization. The basic idea is to construct a tractable sub-problem by considering a randomly drawn finite subset of constraints. The approach has been shown numerically efficient and it has been widely applied to various stochastic and robust programs (see Calafiore and Campi [15]). In a more recent development, Anderson et al. [2] propose a CVaR approximation scheme to a semi-infinite constraint system, and then apply the well-known sample average approximation to the CVaR (of the constraint function); see also Liu and Xu [24] where the approach is applied to mathematical programs with semi-infinite complementarity constraints.

Let $\boldsymbol{\xi}^1, \cdots, \boldsymbol{\xi}^S$ be random variables which are independent and follow identical distribution to that of $\boldsymbol{\xi}$. Let $\mathcal{S} = \{1, \cdots, S\}$. We consider the following discretization problem as an approximation to problem (3.6):

$$
\begin{aligned}
\min_{x,y(\xi^s):s\in\mathcal{S},\alpha_0,\boldsymbol{\alpha}} \quad & c^T x + \alpha_0 + \boldsymbol{\alpha}^T \boldsymbol{\mu} \\
\text{s.t.} \quad & x \in \mathcal{X}, \\
& h^T y(\xi^s) \le \alpha_0 + \boldsymbol{\alpha}^T \xi^s, \text{ for all } s \in \mathcal{S}, \\
& y(\xi^s) \in \mathcal{Y}(x,\xi^s), \text{ for all } s \in \mathcal{S},
\end{aligned}
\tag{3.7}
$$

where ξ^s is a realization of $\boldsymbol{\xi}$ for $s \in \mathcal{S}$. In what follows, we show the convergence of the optimal value of (3.7) to its true counterpart as the sample size increases. To this end, we

consider an equivalent form of (3.7) which is presented in terms of the optimal value function of the second-stage problem:

$$\begin{aligned}
\min_{x,\alpha_0,\boldsymbol{\alpha}} \quad & c^T x + \alpha_0 + \boldsymbol{\alpha}^T \boldsymbol{\mu} \\
\text{s.t.} \quad & x \in \mathcal{X}, \\
& g(x, \xi^s) \leq \alpha_0 + \boldsymbol{\alpha}^T \xi^s, \text{ for all } s \in \mathcal{S}.
\end{aligned} \tag{3.8}$$

A clear benefit of the formulation above is that the decision variables are independent of the sample and this will particularly facilitate the convergence analysis.

To minimize the dependence on the specific details of the objective and constraints functions of problem (3.8) for the convergence analysis and also for the purpose of potential applications of the convergence result, we consider the following general optimization problem

$$\begin{aligned}
\min_{x \in X} \quad & \psi(x) \\
\text{s.t.} \quad & f(x, \xi) \leq 0, \text{ for all } \xi \in \Xi,
\end{aligned} \tag{3.9}$$

where X is a compact set in a finite dimensional space, ψ and f are continuous functions which map from $\mathbb{R}^n$ and $\mathbb{R}^n \times \mathbb{R}^k$ to $\mathbb{R}$ respectively, ξ is the parameter which takes values over a compact set Ξ.

Let $\xi^1, \cdots, \xi^N$ be independent and identically distributed random variables with the same distribution as ξ. We consider the discretized problem

$$\begin{aligned}
\min_{x \in X} \quad & \psi(x) \\
\text{s.t.} \quad & f(x, \xi^i) \leq 0, i = 1, \cdots, N.
\end{aligned} \tag{3.10}$$

Let v and v_N denote, respectively, the optimal values of program (3.9) and program (3.10).

Lemma 3.1 *Assume that (a) ψ is Lipschitz continuous, (b) Ξ is a compact set, (c) ξ is a continuously distributed [2] and there exist positive constants K and τ independent of x, such that for each $x \in X$ there exist $\alpha_0(x) < f^*(x) := \max_{y \in \Xi} f(x, y)$ with*

$$1 - F_x(\alpha) \geq K \left(f^*(x) - \alpha \right)^\tau, \text{ for all } \alpha \in (\alpha_0(x), f^*(x)), \tag{3.11}$$

where F_x denotes the cumulative distribution function of $f(x, \xi)$ [3], and (d) f is Lipschitz continuous in x with integrable Lipschitz modulus (w.r.t. the distribution of ξ). Then for any positive number ϵ, there exist positive constants $C(\epsilon)$ and $A(\epsilon)$ such that

$$\text{Prob}(|v_N - v| \geq \epsilon) \leq C(\epsilon) e^{-A(\epsilon)N} \tag{3.12}$$

for N sufficiently large.

Proof: The thrust of the proof is to use CVaR and its sample average approximation to approximate the semi-infinite constraints of (3.9) which is similar to the convergence analysis

---

[2] Although ξ is a deterministic parameter here, we may regard it as a random variable, and by writing $f(x, \xi) \leq 0$ we mean that, for every realization of ξ, the inequality holds.

[3] Note that ξ could be any random variable which follows a continuous distribution with support set Ξ and the cumulative distribution function satisfying (3.11).

in Anderson et al. [2]. However, there are a few important distinctions: (a) the convergence here is for the randomization scheme (3.10) rather than the sample average approximation of the CVaR approximation of the semi-infinite constraints, (b) the underlying functions in the objective and constraints are not necessarily convex, and (c) the decision vector may consist of some integer variables.

Let

$$\Phi(x) := \sup_{\xi \in \Xi} f(x, \xi) \qquad \text{and} \qquad \Phi_N(x) := \sup_{i=1,\cdots,N} f(x, \xi^i),$$

and let $\mathcal{F}$ and $\mathcal{F}^N$ denote the feasible set of problem (3.9) and problem (3.10) respectively. Then

$$\mathcal{F} = \{x : \Phi(x) = 0\} \qquad \text{and} \qquad \mathcal{F}^N = \{x : \Phi_N(x) = 0\}.$$

Moreover, since $\Phi_N(x) \leq \Phi(x)$, $\mathcal{F} \subset \mathcal{F}^N$. For $\beta \in (0, 1)$, let

$$\text{CVaR}_\beta(f(x, \xi)) := \sup_\eta \left\{ \eta + \frac{1}{1 - \beta} \int_{y \in Y} (f(x, y) - \eta)_+ \rho(y) dy \right\}$$

and

$$\Psi_\beta^N(x) := \sup_\eta \left\{ \eta + \frac{1}{(1 - \beta)N} \sum_{j=1}^N (f(x, \xi^j) - \eta)_+ \right\}$$

where $\rho(\cdot)$ denotes the density function of the random variable ξ, $(a)_+ = \max(0, a)$ for $a \in \mathbb{R}$. In the literature, $\text{CVaR}_\beta(f(x, \xi))$ is known as conditional value at risk and $\Psi_\beta^N(x)$ is its sample average approximation (see Rockafellar and Uryasev [36] and Anderson et al. [2]). It is well known that the maximum w.r.t. η in the above formulation is achieved at a finite η. In other words, we may restrict the maximum to be taken within a closed interval $[-a, a]$ for a sufficiently large, see Rockafellar and Uryasev [36]. It is easy to verify that

$$\Psi_\beta^N(x) \leq \Phi_N(x) \leq \Phi(x). \tag{3.13}$$

We proceed the rest of the proof in four steps.
Step 1. By the definition of CVaR,

$$\text{CVaR}_\beta(f(x, \xi)) \leq \Phi(x)$$

for any $\beta \in (0, 1)$ (see [2]). Moreover, under condition (c), it follows by [2, Theorem 2.1] that

$$|\text{CVaR}_\beta(f(x, \xi)) - \Phi(x)| \leq \frac{1}{K^{1/\tau}} \frac{\tau}{1 + \tau} (1 - \beta)^{1/\tau}. \tag{3.14}$$

Therefore by driving β to 1, we obtain

$$\sup_{x \in X} |\text{CVaR}_\beta(f(x, \xi)) - \Phi(x)| \to 0.$$

Step 2. Using the inequalities (3.13), we have

$$
\begin{aligned}
|\Phi_N(x) - \Phi(x)| &\leq |\Psi_\beta^N(x) - \Phi(x)| \\
&\leq |\Psi_\beta^N(x) - \mathrm{CVaR}_\beta\left(f(x,\xi)\right)| + |\mathrm{CVaR}_\beta\left(f(x,\xi)\right) - \Phi(x)|. \quad (3.15)
\end{aligned}
$$

Let δ be a small positive number. By (3.14), we may set β sufficiently close to 1 such that

$$
\sup_{x \in X} |\mathrm{CVaR}_\beta\left(f(x,\xi)\right) - \Phi(x)| \leq \frac{\delta}{2}.
$$

On the other hand, since Ξ is compact and f is Lipschitz continuous in x with integrable modulus, by virtue of [40, Theorem 5.1], there exist positive constants $C(\delta)$ and $A(\delta)$ such that

$$
\begin{aligned}
&\mathrm{Prob}\left(\sup_{x \in X} |\Psi_\beta^N(x) - \mathrm{CVaR}_\beta\left(f(x,\xi)\right)| \geq \delta/2 \right) \\
&\leq \mathrm{Prob}\left(\frac{1}{1-\beta} \sup_{x \in X} \sup_{\eta \in [-a,a]} \left| \frac{1}{N} \sum_{j=1}^N (f(x,\xi^j) - \eta)_+ - \mathbb{E}_P[(\eta - f(x,\xi))_+] \right| \geq \delta/2 \right) \\
&\leq C(\delta) e^{-A(\delta)N}
\end{aligned}
$$

when N is sufficiently large. Here, in the first inequality, we are using the fact that the maximum w.r.t. η is achieved in $[-a,a]$ for some appropriate positive constant a; see similar discussions in Xu and Zhang [50]. Therefore,

$$
\begin{aligned}
\mathrm{Prob}\left(\sup_{x \in X} |\Phi_N(x) - \Phi(x)| \geq \delta \right) &\leq \mathrm{Prob}\left(\sup_{x \in X} |\mathrm{CVaR}_\beta^N\left(f(x,\xi)\right) - \mathrm{CVaR}_\beta\left(f(x,\xi)\right)| \geq \delta/2 \right) \\
&\leq C(\delta) e^{-A(\delta)N}. \quad (3.16)
\end{aligned}
$$

Step 3. For small positive number t, let

$$
R(t) = \min_x \left\{ \Phi(x) : d(x, \mathcal{F}) \right\} \geq t,
$$

where $d(x, \mathcal{F})$ denotes the distance from point x to set $\mathcal{F}$. Obviously $R(t) > 0$, it is monotonically increasing, and $R(t) \to 0$ as $t \downarrow 0$. Therefore, it is easy to observe that

$$
d(x, \mathcal{F}) \geq t \iff \Phi(x) \geq R(t).
$$

Let

$$
\mathbb{D}(\mathcal{F}^N, \mathcal{F}) := \sup_{x \in \mathcal{F}^N} d(x, \mathcal{F})
$$

and

$$
\mathbb{H}(\mathcal{F}^N, \mathcal{F}) := \max\left(\mathbb{D}(\mathcal{F}^N, \mathcal{F}), \mathbb{D}(\mathcal{F}, \mathcal{F}^N) \right),
$$

where $\mathbb{H}$ is the Hausdorff distance between $\mathcal{F}^N$ and $\mathcal{F}$. Since both $\mathcal{F}^N$ and $\mathcal{F}$ are bounded, the Hausdorff distance is well defined. Moreover, since $\mathcal{F} \subset \mathcal{F}^N$, $\mathbb{H}(\mathcal{F}^N, \mathcal{F}) = \mathbb{D}(\mathcal{F}^N, \mathcal{F})$. In what follows, we estimate $\mathrm{Prob}\left(\mathbb{D}(\mathcal{F}^N, \mathcal{F}) \geq \delta \right)$. For any $x_N \in F^N$, since $\Phi_N(x_N) = 0$, then

$$
\Phi(x_N) \geq R(t) \iff \Phi(x_N) - \Phi_N(x_N) \geq R(t).
$$

13

Therefore,

$$\text{Prob}\Big(\mathbb{D}(\mathcal{F}^N, \mathcal{F}) \geq t\Big) \leq \text{Prob}\left(\sup_{x \in \mathcal{F}^N} |\Phi(x)| \geq R(t)\right)$$

$$\leq \text{Prob}\left(\sup_{x \in X} |\Phi(x) - \Phi_N(x)| \geq R(t)\right). \qquad (3.17)$$

Step 4. Let $x^* \in \mathcal{F}$ and $x^N \in \mathcal{F}^N$ be the optimal solutions to (3.9) and (3.10). Then, by the Lipschitz continuity of ψ,

$$|v^N - v^*| = |\psi(x^N) - \psi(x^*)| \leq L\|x^N - x^*\| \leq L\mathbb{H}(\mathcal{F}^N, \mathcal{F})$$

where L denotes the Lipschitz modulus of ψ. By (3.17), we deduce

$$\text{Prob}(|v^N - v^*| \geq \epsilon) \leq \text{Prob}\Big(\mathbb{H}(\mathcal{F}^N, \mathcal{F}) \geq \epsilon/L\Big) \leq \text{Prob}\left(\sup_{x \in X} |\Phi(x) - \Phi_N(x)| \geq R(\epsilon/L)\right).$$

The rest follows from (3.16) with $\delta = R(\epsilon/L)$. The proof is complete. $\qquad \square$

With Lemma 3.1, we are ready to state the convergence of problem (3.8).

Theorem 3.1 *Let ϑ and ϑ_N denote the optimal value of (SIP-UC) and (3.8) respectively. Assume that ξ follows a uniform distribution [4]. Then for any positive number ϵ, there exists positive constants $C(\epsilon)$ and $\beta(\epsilon)$ such that*

$$\text{Prob}(|\vartheta_N - \vartheta| \geq \epsilon) \leq C(\epsilon)e^{-\beta(\epsilon)N} \qquad (3.18)$$

for N sufficiently large.

Proof: It suffices to verify the conditions of Lemma 3.1. Conditions (a) and (b) are obvious since the objective function is linear and Ξ is compact problem (3.8). Condition (c) is satisfied because $g(x, \xi)$ also follows a uniform distribution for each fixed x and the cumulative distribution function of $g(x, \xi)$ is a linear function. Let us verify condition (d). Following Walkup et al. [45] and Nožička [28] (see also Liu et al. [25, Lemma 4.3]) that $g(x, \xi)$ is Lipschitz continuous w.r.t. x and ξ and since Ξ is compact, $g(x, \xi)$ is Lipschitz continuous in x with integrable Lipschitz modulus. The proof is complete. $\qquad \square$

In Theorem 3.1, we assume that ξ follows a uniform distribution. It might be interesting to show the conclusion when ξ follows a general continuous distribution with positive density function in the interior of Ξ; that is, $g(x, \xi)$ satisfies (3.11). We leave this for our future work.

The detailed formulation for the sample approximation of (SIP-UC) problem is given by

---

[4] Following the comments at a footnote of Lemma 3.1, the distribution of ξ can be any so long as its support set coincides with Ξ and it satisfies consistent tail condition (3.14) because the distribution is only used for generating samples of the random discretization scheme. We make it easier by considering uniform distribution which obviously satisfies the tail condition as the density is a postive constant.

the mixed integer linear program below,

$$\min_{\mathbf{u},\mathbf{q},\mathbf{r},\hat{\mathbf{r}}(.),\boldsymbol{\alpha}} \quad c_i^f u_{it} + c_i^l q_{it} + c_i^{r,up} r_{it}^{up} + c_i^{r,dw} r_{it}^{dw} + \alpha_0 + \sum_{t=1}^{T} \alpha_t \mu_t$$

$$
\begin{aligned}
\text{s.t.} \quad & q_{it} + r_{it}^{up} \leq \overline{q}_i u_{it}, && \forall t, i, \\
& q_{it} - r_{it}^{dw} \geq \underline{q}_i u_{it}, && \forall t, i, \\
& \underline{r}_i^{up} \leq r_{it}^{up} \leq \overline{r}_i^{up}, && \forall t, i, \\
& \underline{r}_i^{dw} \leq r_{it}^{dw} \leq \overline{r}_i^{dw}, && \forall t, i, \\
& q_{it}, r_{it}^{dw}, r_{it}^{up} \in \mathbb{R}^+, \; u_{it} \in \{0,1\}, && \forall t, i, \\
& \sum_i \sum_t \{\hat{c}_i^{up} \hat{r}_{it}^{up}(\xi^s) + \hat{c}_i^{dw} \hat{r}_{it}^{dw}(\xi^s)\} + \sum_t \{c_t^{ls} S_t(\xi^s) + c_t^{ws} W_t(\xi^s)\} \leq
\end{aligned}
$$

$$
\begin{aligned}
& \alpha_0 + \sum_{t=1}^{T} \alpha_t \xi_t^s, && \forall s, \\
& \sum_i \hat{q}_{it}(\xi^s) + S_t(\xi^s) - W_t(\xi^s) = l_t(\xi^s), && \forall t, s, \\
& 0 \leq \hat{r}_{it}^{up}(\xi^s) \leq r_{it}^{up}, && \forall t, i, s, \\
& 0 \leq \hat{r}_{it}^{dw}(\xi^s) \leq r_{it}^{dw}, && \forall t, i, s, \\
& Q_t - (q_{it} + r_{it}^{up}) + S_t(\underline{\xi}) \geq l_t(\underline{\xi}), && \forall t, i, \\
& \hat{q}_{it}(\xi) = q_{it} + \hat{r}_{it}^{up}(\xi) - \hat{r}_{it}^{dw}(\xi), && \forall t, i, \\
& \hat{q}_{it}(\xi^s) - \hat{q}_{i(t-1)}(\xi^s) \leq RU_i u_{i(t-1)} + SU_i(1 - u_{i(t-1)}), && \forall t, i, s, \\
& \hat{q}_{i(t-1)}(\xi^s) - \hat{q}_{it}(\xi^s) \leq RD_i u_{it} + SD_i(1 - u_{it}), && \forall t, i, s,
\end{aligned}
$$

where the first five constraints are the first-stage constraints, and the last six constraints represent the remaining constraints in problem (3.7).

4 Mixture distribution approach

In the absence of complete information on the underlying distribution of random variables, the decision maker could integrate information obtained through various channels to construct a mixture probability distribution. The use of mixture distribution in the context of robust optimization could be traced back to Peel and McLachlan [31] and more recently in Zhu and Fukushima [51] for portfolio optimization problems.

To define the ambiguity set corresponding to the robust problem (2.13), let $P_j, j = 1, \cdots, L$ be a set of probability measures such that $\mathbb{E}_{P_j}[g(x, \boldsymbol{\xi})]$ is well defined for $j = 1, \cdots, L$. The ambiguity set under mixture distribution can then be defined as follows,

$$\mathcal{P} := \Big\{ \sum_{j=1}^{L} \gamma_j P_j : \quad \sum_{j=1}^{L} \gamma_j = 1, \quad \gamma_j \geq 0, \; \forall j = 1, \cdots, L \Big\},$$

where γ_j denotes the weight of distribution j. The probability distributions $P_1, \cdots, P_L$ are assumed to be known and true probability distribution is assumed to be in their convex hull. For any realization of the distribution $P := \sum_j \gamma_j P_j$ we have

$$\mathbb{E}_P[g(x, \boldsymbol{\xi})] = \sum_{j=1}^{L} \gamma_j \mathbb{E}_{P_j}[g(x, \boldsymbol{\xi})].$$

Under the mixture distribution, the inner maximization problem in (2.13) can then be rewritten as follows,

$$
\begin{aligned}
H(x) = \ \sup_{\gamma} \ & \sum_{j=1}^{L} \gamma_j \mathbb{E}_{P_j}[g(x, \boldsymbol{\xi})] \\
\text{s.t.} \ & \sum_{j} \gamma_j = 1, \\
& \gamma_j \geq 0, \qquad \forall j = 1, \cdots, L.
\end{aligned}
\tag{4.1}
$$

Let λ be the dual variable corresponding to the first constraint in (4.1). Then the dual of the above problem can be written as

$$
\begin{aligned}
H_D(x) = \ \min_{\lambda \in \mathbb{R}} \ & \lambda \\
\text{s.t.} \ & \mathbb{E}_{P_j}[g(x, \boldsymbol{\xi})] \leq \lambda, \quad \forall j = 1, \cdots, L.
\end{aligned}
\tag{4.2}
$$

Proposition 4.1 *In the case when $\boldsymbol{\xi}$ has a finite discrete distribution with the support set containing a finite number of values $\boldsymbol{\xi}^1, \cdots, \boldsymbol{\xi}^N$ and for a given first-stage decision x, model (4.2) is equivalent to*

$$
\begin{aligned}
\min_{y^1, \cdots, y^N, \lambda} \ & \lambda \\
\text{s.t.} \ & \sum_{k=1}^{N} p_j^k (h^T y^k) \leq \lambda, \forall j = 1, \cdots, L, \\
& y^k \in \mathcal{Y}(x, \boldsymbol{\xi}^k), \forall k = 1, \cdots, N,
\end{aligned}
\tag{4.3}
$$

where p_j^k is the probability measure of P_j in scenario k.

Proof: If $\boldsymbol{\xi}$ has a discrete distribution with a finite number of scenarios $\boldsymbol{\xi}^1, \cdots, \boldsymbol{\xi}^N$, then model (4.2) can be written as

$$
\begin{aligned}
\min_{\lambda} \ & \lambda \\
\text{s.t.} \ & \sum_{k=1}^{N} p_j^k g(x, \boldsymbol{\xi}^k) \leq \lambda, \quad \forall j = 1, \cdots, L,
\end{aligned}
$$

where $g(x, \boldsymbol{\xi}^k)$ refers to the second-stage problem

$$
\begin{aligned}
g(x, \boldsymbol{\xi}^k) = \ \min_{y} \ & h^T y \\
\text{s.t.} \ & y \in \mathcal{Y}(x, \boldsymbol{\xi}^k).
\end{aligned}
$$

Let us denote the optimal solution of the above problem as

$$
\hat{y}^k \ = \ \underset{y \in \mathcal{Y}(x, \boldsymbol{\xi}^k)}{\arg\min} h^T y, \forall k.
$$

Then we have $g(x, \boldsymbol{\xi}^k) = h^T \hat{y}^k$. Let us define $\hat{\lambda} = \sum_{k=1}^{N} p_j^k (h^T \hat{y}^k)$. It is clear that $(\hat{\lambda}, \hat{y})$ is a feasible solution to problem (4.3). Let $(\tilde{\lambda}, \tilde{y}^k : k = 1, \cdots, N)$ be an optimal solution to

problem(4.3). Then $h^T \tilde{y}^k \geq h^T \hat{y}^k$ for all k by the definition of $\hat{y}^k$. Thus, we have

$$\sum_{k=1}^{N} p_j^k (h^T \tilde{y}^k) \; \geq \; \sum_{k=1}^{N} p_j^k (h^T \hat{y}^k), \quad \forall j.$$

Therefore

$$\min_j \left\{ \sum_{k=1}^{N} p_j^k (h^T \tilde{y}^k) \right\} \; \geq \; \min_j \left\{ \sum_{k=1}^{N} p_j^k (h^T \hat{y}^k) \right\}$$

and $\tilde{\lambda} \geq \hat{\lambda}$. Since $(\hat{\lambda}, \hat{y})$ is a feasible solution of (4.3), it must also be an optimal solution. $\square$

Thus, when $\boldsymbol{\xi}$ has a finite discrete distribution, the original robust problem (2.13) can be rewritten as

$$
\begin{aligned}
\text{(Mix-UC)} \quad & \min_{x,y(.),\lambda} && c^T x + \lambda \\
& \text{s.t.} && x \in \mathcal{X}, \\
& && \sum_{k=1}^{N} p_j^k (h^T y^k) \leq \lambda, \forall j = 1, \cdots, L, \\
& && y^k \in \mathcal{Y}(x, \boldsymbol{\xi}^k), \forall k = 1, \cdots, N.
\end{aligned}
\tag{4.4}
$$

Note that when P_j follows a continuous distribution, it might be difficult to compute the expected value of the functions in the constraints of (4.2). A well-known approach to resolving this issue is to use sample average approximation (SAA). For a fixed j, let $\boldsymbol{\xi}_j^1, \cdots, \boldsymbol{\xi}_j^{N_j}$ denote independent and identically random sampling of $\boldsymbol{\xi}$. Then $\mathbb{E}_{p_j}[g(x, \boldsymbol{\xi})]$ can be approximated by

$$\frac{1}{N_j} \sum_{k=1}^{N_j} g(x, \boldsymbol{\xi}_j^k).$$

Consequently, problem (4.4) can be approximated by

$$
\begin{aligned}
& \min_{x,y(.),\lambda} && c^T x + \lambda \\
& \text{s.t.} && x \in \mathcal{X}, \\
& && \frac{1}{N_j} \sum_{k=1}^{N_j} h^T y_j^k \leq \lambda, \forall j = 1, \cdots, L, \\
& && y_j^k \in \mathcal{Y}(x, \boldsymbol{\xi}_j^k), \forall k = 1, \cdots, N_j, j = 1, \cdots, L.
\end{aligned}
\tag{4.5}
$$

The overall SAA of the Mix-UC problem can be written as follows,

$$\min_{\mathbf{u},\mathbf{q},\mathbf{r},\hat{\mathbf{r}},\mathbf{W},\mathbf{S},\lambda} \quad c_i^f u_{it} + c_i^l q_{it} + c_i^{r,up} r_{it}^{up} + c_i^{r,dw} r_{it}^{dw} + \lambda$$

$$\text{s.t.} \quad q_{it} + r_{it}^{up} \leq \bar{q}_i u_{it}, \qquad \forall t, i,$$

$$q_{it} - r_{it}^{dw} \geq \underline{q}_i u_{it}, \qquad \forall t, i,$$

$$\underline{r}_i^{up} \leq r_{it}^{up} \leq \bar{r}_i^{up}, \qquad \forall t, i,$$

$$\underline{r}_i^{dw} \leq r_{it}^{dw} \leq \bar{r}_i^{dw}, \qquad \forall t, i,$$

$$q_{it}, r_{it}^{dw}, r_{it}^{up} \in \mathbb{R}^+, \ u_{it} \in \{0, 1\}, \qquad \forall t, i,$$

$$\frac{1}{N_j} \sum_{k=1}^{N_j} \left[\sum_i \sum_t \{ \hat{c}_i^{up} \hat{r}_{it}^{up}(\xi_j^k) + \right.$$

$$\left. \hat{c}_i^{dw} \hat{r}_{it}^{dw}(\xi_j^k) \} + \sum_t \{ c_t^{ls} S_t(\xi_j^k) + c_t^{ws} W_t(\xi_j^k) \} \right] \leq \lambda \quad \forall j,$$

$$\sum_i \hat{q}_{it}(\xi_j^k) + S_t(\xi_j^k) - W_t(\xi_j^k) = l_t(\xi_j^k) \qquad \forall t, k, j,$$

$$0 \leq \hat{r}_{it}^{up}(\xi_j^k) \leq r_{it}^{up} \qquad \forall t, i, k, j,$$

$$0 \leq \hat{r}_{it}^{dw}(\xi_j^k) \leq r_{it}^{dw} \qquad \forall t, i, k, j,$$

$$Q_t - (q_{it} + r_{it}^{up}) + S_t(\underline{\xi}) \geq l_t(\underline{\xi}), \qquad \forall t, i,$$

$$\hat{q}_{it}(\xi) = q_{it} + \hat{r}_{it}^{up}(\xi) - \hat{r}_{it}^{dw}(\xi), \qquad \forall t, i,$$

$$\hat{q}_{it}(\xi_j^k) - \hat{q}_{i(t-1)}(\xi_j^k) \leq RU_i u_{i(t-1)} + SU_i(1 - u_{i(t-1)}), \quad \forall t, i, k, j,$$

$$\hat{q}_{i(t-1)}(\xi_j^k) - \hat{q}_{it}(\xi_j^k) \leq RD_i u_{it} + SD_i(1 - u_{it}), \qquad \forall t, i, k, j.$$

Note that the SAA approach for Mix-UC results in a mixed integer linear program (MILP).

5 Case study

In this section, we carry out some numerical experiments to evaluate the proposed mathematical methods and numerical schemes. To facilitate further exposition and reading, we develop a list of models in Table 2.

Table 2: Reference of the methodologies

Abbreviations	Problem	Method	Known information on uncertainty
Sto-UC	(2.1)-(2.12)	Stochastic UC	Probability distribution (P)
SIP-UC	(3.5)	Robust UC with semi-infinite formulation	First moments (μ)
Mix-UC	(4.4)	Robust UC with mixture distribution	A set of known distributions

5.1 Data

We consider an illustrative case study based on a system with 10 generating units. The data for the generators are based on work by Pozo and Contreras [32] and includes the cost and limitation of the generation and reserve utilization as summarised in Table 3.

Table 3: Ten-unit system: generation and reserve cost and capacities

| Unit | Costs [$/MW] | | | | | Capacities [MW] | | |
| | Fixed [$] cost | Variable cost | Reserve up/down (scheduled) | Reserve up (actual) | Reserve down (actual) | Minimum output | Maximum output | Reserve up/down |
i	c_i^f	c_i^l	$c_i^{r,up}$ / $c_i^{r,dw}$	$\hat{c}_i^{up}$	$\hat{c}_i^{dw}$	$\underline{q}_i$	$\overline{q}_i$	$\overline{r}_i^{up}$ / $\overline{r}_i^{dw}$
1	2,550	16.19	1.80	17.81	-14.57	150	455	153
2	2,550	17.26	1.92	18.99	-15.53	150	455	153
3	1,300	16.60	1.84	18.26	-14.94	70	180	55
4	1,300	16.50	1.83	18.15	-14.85	70	180	55
5	1,620	19.70	2.19	21.67	-17.73	50	165	58
6	800	22.26	2.47	24.49	-20.03	30	90	30
7	850	27.74	3.08	30.51	-24.97	40	85	23
8	550	25.92	2.88	28.51	-23.33	20	60	20
9	550	27.27	3.03	30.00	-24.54	20	60	20
10	550	27.79	3.09	30.57	-25.01	20	60	20

Stochastic base case

The hourly power demand is assumed to be known, and under the two-stage stochastic setting (Sto-UC) the wind output is assumed to follow a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with mean $\boldsymbol{\mu}$, standard deviation σ, correlation C and covariance $\Sigma = \sigma C \sigma^T$. The hourly demand, wind output mean values $\boldsymbol{\mu}$, and the standard deviations σ for a 24-hour period are given in Table 4. The corresponding mean values for the stochastic net load are also presented in Figure 1.

Table 4: Hourly demand and mean/standard deviation of the wind

| Hour | Demand | Wind | | Hour | Demand | Wind | |
t	d_t	Mean (μ_t)	SD(σ_t)	t	d_t	Mean (μ_t)	SD(σ_t)
1	1127	282	42.3	13	2254	564	126.8
2	1208	302	47.2	14	2093	523	121
3	1369	342	55.6	15	1932	483	114.7
4	1530	383	64.5	16	1691	423	103
5	1610	403	70.4	17	1610	403	100.6
6	1771	443	80.2	18	1771	443	113.5
7	1852	463	86.8	19	1932	483	126.8
8	1932	483	93.6	20	2254	564	151.4
9	2093	523	104.7	21	2093	523	143.9
10	2254	564	116.2	22	1771	443	124.5
11	2335	584	124	23	1449	362	104.1
12	2415	604	132.1	24	1288	322	94.6
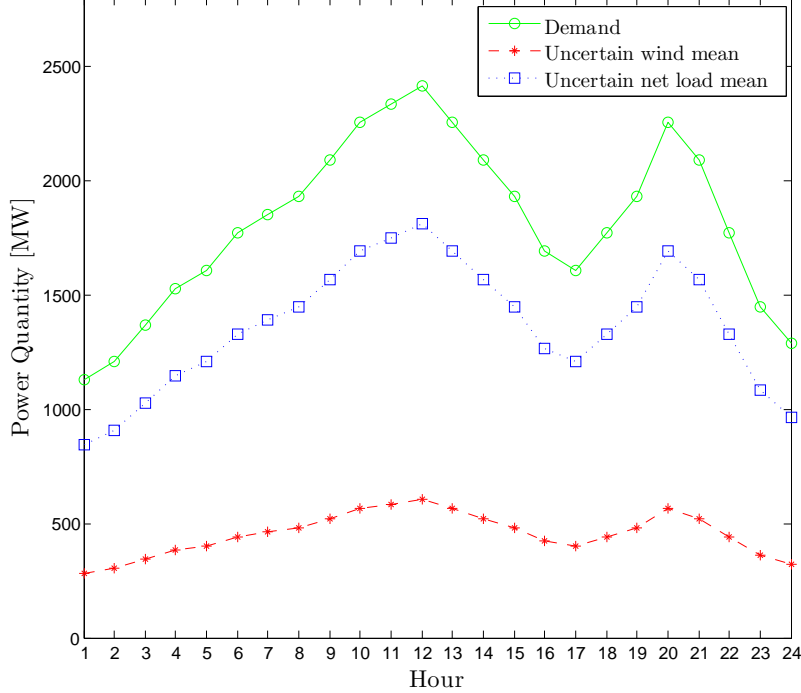
Figure 1: Nominal values of demand and mean wind

The correlation matrix C is based on the hourly Danish wind output data[5] for the year 2013 and is given by

$$
C = \begin{pmatrix}
1 & 0.994 & 0.979 & 0.962 & 0.942 & 0.913 & 0.88 & 0.844 & 0.812 & 0.778 & 0.744 & 0.705 & 0.665 & 0.629 & 0.597 & 0.566 & 0.54 & 0.51 & 0.479 & 0.462 & 0.449 & 0.428 & 0.399 & 0.372 \\
0.994 & 1 & 0.994 & 0.982 & 0.965 & 0.937 & 0.905 & 0.87 & 0.838 & 0.806 & 0.774 & 0.735 & 0.697 & 0.66 & 0.627 & 0.595 & 0.568 & 0.535 & 0.504 & 0.486 & 0.471 & 0.449 & 0.42 & 0.391 \\
0.979 & 0.994 & 1 & 0.995 & 0.982 & 0.957 & 0.927 & 0.893 & 0.862 & 0.829 & 0.797 & 0.759 & 0.719 & 0.682 & 0.648 & 0.616 & 0.587 & 0.553 & 0.522 & 0.505 & 0.489 & 0.467 & 0.439 & 0.411 \\
0.962 & 0.982 & 0.995 & 1 & 0.994 & 0.975 & 0.949 & 0.919 & 0.888 & 0.856 & 0.823 & 0.784 & 0.744 & 0.706 & 0.672 & 0.64 & 0.611 & 0.576 & 0.544 & 0.527 & 0.511 & 0.487 & 0.459 & 0.43 \\
0.942 & 0.965 & 0.982 & 0.994 & 1 & 0.992 & 0.974 & 0.949 & 0.922 & 0.892 & 0.86 & 0.821 & 0.781 & 0.741 & 0.705 & 0.673 & 0.643 & 0.608 & 0.576 & 0.558 & 0.54 & 0.515 & 0.486 & 0.454 \\
0.913 & 0.937 & 0.957 & 0.975 & 0.992 & 1 & 0.993 & 0.976 & 0.953 & 0.926 & 0.895 & 0.858 & 0.817 & 0.776 & 0.74 & 0.707 & 0.677 & 0.641 & 0.609 & 0.59 & 0.57 & 0.542 & 0.511 & 0.477 \\
0.88 & 0.905 & 0.927 & 0.949 & 0.974 & 0.993 & 1 & 0.993 & 0.976 & 0.953 & 0.924 & 0.888 & 0.847 & 0.805 & 0.768 & 0.735 & 0.705 & 0.669 & 0.636 & 0.616 & 0.594 & 0.563 & 0.53 & 0.494 \\
0.844 & 0.87 & 0.893 & 0.919 & 0.949 & 0.976 & 0.993 & 1 & 0.992 & 0.974 & 0.947 & 0.913 & 0.873 & 0.831 & 0.795 & 0.762 & 0.731 & 0.696 & 0.663 & 0.642 & 0.619 & 0.586 & 0.55 & 0.511 \\
0.812 & 0.838 & 0.862 & 0.888 & 0.922 & 0.953 & 0.976 & 0.992 & 1 & 0.992 & 0.969 & 0.939 & 0.902 & 0.862 & 0.827 & 0.796 & 0.766 & 0.731 & 0.699 & 0.678 & 0.652 & 0.616 & 0.577 & 0.536 \\
0.778 & 0.806 & 0.829 & 0.856 & 0.892 & 0.926 & 0.953 & 0.974 & 0.992 & 1 & 0.99 & 0.967 & 0.935 & 0.899 & 0.866 & 0.836 & 0.806 & 0.77 & 0.735 & 0.711 & 0.683 & 0.644 & 0.604 & 0.562 \\
0.744 & 0.774 & 0.797 & 0.823 & 0.86 & 0.895 & 0.924 & 0.947 & 0.969 & 0.99 & 1 & 0.991 & 0.969 & 0.94 & 0.91 & 0.882 & 0.854 & 0.819 & 0.78 & 0.754 & 0.723 & 0.682 & 0.641 & 0.598 \\
0.705 & 0.735 & 0.759 & 0.784 & 0.821 & 0.858 & 0.888 & 0.913 & 0.939 & 0.967 & 0.991 & 1 & 0.991 & 0.971 & 0.946 & 0.922 & 0.895 & 0.861 & 0.822 & 0.794 & 0.761 & 0.719 & 0.677 & 0.635 \\
0.665 & 0.697 & 0.719 & 0.744 & 0.781 & 0.817 & 0.847 & 0.873 & 0.902 & 0.935 & 0.969 & 0.991 & 1 & 0.993 & 0.976 & 0.957 & 0.933 & 0.901 & 0.863 & 0.835 & 0.801 & 0.759 & 0.719 & 0.679 \\
0.629 & 0.66 & 0.682 & 0.706 & 0.741 & 0.776 & 0.805 & 0.831 & 0.862 & 0.899 & 0.94 & 0.971 & 0.993 & 1 & 0.994 & 0.98 & 0.961 & 0.933 & 0.899 & 0.871 & 0.837 & 0.796 & 0.757 & 0.719 \\
0.597 & 0.627 & 0.648 & 0.672 & 0.705 & 0.74 & 0.768 & 0.795 & 0.827 & 0.866 & 0.91 & 0.946 & 0.976 & 0.994 & 1 & 0.995 & 0.981 & 0.958 & 0.928 & 0.901 & 0.867 & 0.827 & 0.789 & 0.753 \\
0.566 & 0.595 & 0.616 & 0.64 & 0.673 & 0.707 & 0.735 & 0.762 & 0.796 & 0.836 & 0.882 & 0.922 & 0.957 & 0.98 & 0.995 & 1 & 0.994 & 0.977 & 0.952 & 0.927 & 0.894 & 0.855 & 0.818 & 0.783 \\
0.54 & 0.568 & 0.587 & 0.611 & 0.643 & 0.677 & 0.705 & 0.731 & 0.766 & 0.806 & 0.854 & 0.895 & 0.933 & 0.961 & 0.981 & 0.994 & 1 & 0.993 & 0.974 & 0.951 & 0.92 & 0.882 & 0.847 & 0.814 \\
0.51 & 0.535 & 0.553 & 0.576 & 0.608 & 0.641 & 0.669 & 0.696 & 0.731 & 0.77 & 0.819 & 0.861 & 0.901 & 0.933 & 0.958 & 0.977 & 0.993 & 1 & 0.991 & 0.973 & 0.944 & 0.908 & 0.875 & 0.845 \\
0.479 & 0.504 & 0.522 & 0.544 & 0.576 & 0.609 & 0.636 & 0.663 & 0.699 & 0.735 & 0.78 & 0.822 & 0.863 & 0.899 & 0.928 & 0.952 & 0.974 & 0.991 & 1 & 0.992 & 0.969 & 0.937 & 0.905 & 0.876 \\
0.462 & 0.486 & 0.505 & 0.527 & 0.558 & 0.59 & 0.616 & 0.642 & 0.678 & 0.711 & 0.754 & 0.794 & 0.835 & 0.871 & 0.901 & 0.927 & 0.951 & 0.973 & 0.992 & 1 & 0.989 & 0.964 & 0.935 & 0.906 \\
0.449 & 0.471 & 0.489 & 0.511 & 0.54 & 0.57 & 0.594 & 0.619 & 0.652 & 0.683 & 0.723 & 0.761 & 0.801 & 0.837 & 0.867 & 0.894 & 0.92 & 0.944 & 0.969 & 0.989 & 1 & 0.99 & 0.969 & 0.944 \\
0.428 & 0.449 & 0.467 & 0.487 & 0.515 & 0.542 & 0.563 & 0.586 & 0.616 & 0.644 & 0.682 & 0.719 & 0.759 & 0.796 & 0.827 & 0.855 & 0.882 & 0.908 & 0.937 & 0.964 & 0.99 & 1 & 0.992 & 0.975 \\
0.399 & 0.42 & 0.439 & 0.459 & 0.486 & 0.511 & 0.53 & 0.55 & 0.577 & 0.604 & 0.641 & 0.677 & 0.719 & 0.757 & 0.789 & 0.818 & 0.847 & 0.875 & 0.905 & 0.935 & 0.969 & 0.992 & 1 & 0.992 \\
0.372 & 0.391 & 0.411 & 0.43 & 0.454 & 0.477 & 0.494 & 0.511 & 0.536 & 0.562 & 0.598 & 0.635 & 0.679 & 0.719 & 0.753 & 0.783 & 0.814 & 0.845 & 0.876 & 0.906 & 0.944 & 0.975 & 0.992 & 1
\end{pmatrix}
$$

SIP formulation

In the SIP formulation, we assume that only the first moment of the uncertain wind power is known and is given by $\boldsymbol{\mu}$. Based on this assumption, we implement the first proposed model and formulate the problem as a robust SIP (SIP-UC). In solving the SIP problem, we limit

---

[5]Available online at `http://energinet.dk`

the support set of random hourly wind to 150 values generated from a multivariate uniform distribution.

Mixture distribution formulation

In the second approach of mixture distribution, we consider a case where the information on the uncertain net load is received through various resources, e.g. advice from a group of experts. While each alternative net load model provides a specific distribution and parameters such as the mean and the covariance, there is no consensus among the decision makers on which model contains the true distribution. Therefore, instead of relying on a particular expert model, we use the mixture model (Mix-UC) to combine all these potential distributions. Specifically, we assume that we are given three different distributions for the stochastic net load, shown in Table 5. We assume that three distributions are equally relevant and therefore they have equal weights in the construction of the mixture distribution. We construct the uncertainty set by drawing 50 random samples from each distribution.

Table 5: Potential distributions for wind power

Distribution	Mean	Covariance	Weight
Multivariate Normal	$0.8\boldsymbol{\mu}$	Σ	$\frac{1}{3}$
Multivariate Normal	$1.2\boldsymbol{\mu}$	Σ	$\frac{1}{3}$
Multivariate Uniform	$\boldsymbol{\mu}$	Σ	$\frac{1}{3}$

5.2   Numerical results

We compare the solutions produced by three different models: (a) a two-stage stochastic UC that uses the known distribution of the random supply as described in Section 2.1, (b) the robust UC that uses only information on the mean value of the random supply as described in Section 3, and (c) the robust UC that 'knows' only a convex hull of a finite set of known distributions containing the true probability distribution as described in Section 4. Table 6 shows the size of each problem as well as the running time for solving each problem. The models have been tested on an MacBook Air Intel Core i5 processor at 1.6 GHz and 4 GB of RAM memory using CPLEX 12.6.1 under MATLAB R2014a.

Table 6: Problems size and computational times

| Model	Scenarios (#)	Constraints (#)	Total variables (#)	Binary variables (#)	CPU time (s)
Sto-UC	150	148,320	80,160	240	539.16
SIP-UC	150	148,470	80,185	240	1894.1
Mix-UC	150	148,322	80,161	240	196.70

The sizes of the three models are dependent on the number of scenarios generated and are very similar to each other in this case study. We generate 150 scenarios for each of the models so that the computational time is reasonable. Our main interest lies in assessing the quality of the solutions proposed by the stochastic model and their robust counterparts.

The cumulative first-stage decisions as well as the second-stage decisions and the total costs for each solution are presented in Table 7. It can be observed that robust solutions result in higher expected costs than the stochastic solution in both the first stage and the second stage. This is contributed to mostly by the increase in the first-stage generation to hedge against the ambiguity of the distribution of the underlying net load uncertainty. The increase of the expected costs makes sense because the stochastic solution was produced by using the truth distribution of the net load; i.e. the truth information is known, while the robust solutions only have the uncertainty set containing the truth distribution. Note that, in practice, we often do not have the truth distribution and hence the stochastic solution must use some 'guess' distribution in the uncertainty set. In that case, the resulting expected cost might be higher than those suggested by the robust solutions as we will demonstrate in the sensitivity analysis later on.

Table 7: Stochastic versus robust solutions

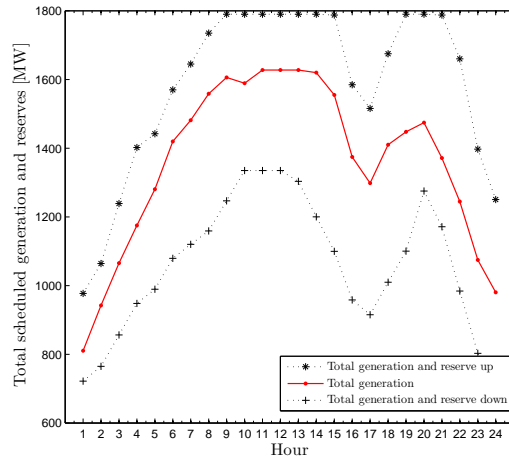| Model | First-stage decisions [Total MW] | | | Cost [$] | | |
	Generation $\sum_t \sum_i q_{it}$	Reserve up $\sum_t \sum_i r_{it}^{up}$	Reserve down $\sum_t \sum_i r_{it}^{dw}$	First stage	Second stage	Total expected
Sto-UC	32,098	3,701	3,882	821,260	50,870	872,130
Mix-UC	33,633	2,910	2,390	854,660	118,270	972,940
SIP-UC	32,661	5,391	7,231	882,740	195,930	1,078,680

The hourly cumulative first-stage generation and reserve levels are presented in Figure 2. It can be observed that, in the case of the Mix-UC solution, the hourly pattern of total generation is generally higher than the base Sto-UC solution. Additionally, the SIP-UC solution provides a greater flexibility for the second stage through a higher level of up and down reserves. It can also be observed that the hourly patterns of the generation and the reserve schedule are similar to the net load quantities, and at the peak hours the maximum generation capacity of generators are scheduled either in term of generation (Mix-UC) or generation with reserve (Sto-UC and SIP-UC).
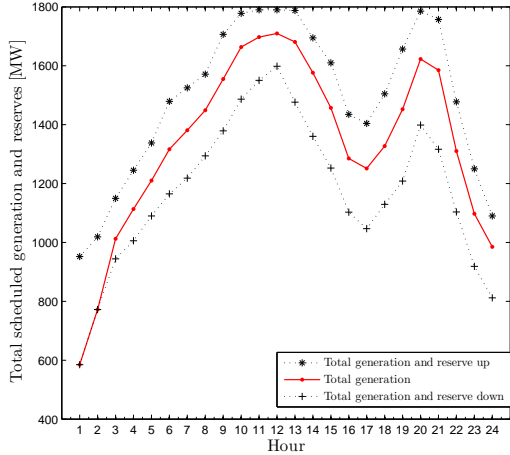
(a) Sto-UC
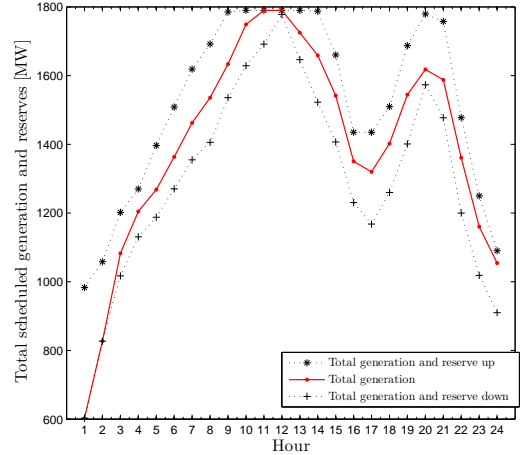
(b) Mix-UC

(c) SIP-UC

Figure 2: UC first-stage solutions

Sensitivity of solutions to variation of the wind distribution

The first-stage decisions of the UC models determines the flexibility of the solution to changes in the actual realization of the wind distribution in the second stage, i.e. we expect the robust solutions to provide greater flexibility if the actual distribution of the uncertainty was different from the assumed distribution. To compare the performance of the Sto-UC model, MIX-UC model, and SIP-UC model, we analyze the effect of deviation of the distribution parameters such as the mean and the covariance matrix from the nominal values. In doing so, we consider a two-stage stochastic structure for each instance, in which the first stage unit commitment and reserve schedules are fixed as the first-stage solutions of the Sto-UC model, MIX-UC model, and SIP-UC model. In each instance, we then solve the second-stage problem using a distribution with different means or covariances for the uncertain wind.

We first study the sensitivity of solution to changes in the actual mean of the distribution. We compare the Sto-UC, Mix-UC, and SIP-UC for instances with the following wind distributions: $\mathcal{N}(a\boldsymbol{\mu}, \Sigma), a = \{0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5\}$ bounded away from

zero (i.e. truncated normal distributions). For each choice of a, we carry out 100 independent runs, each generating 150 i.i.d samples from the corresponding distribution, and use these i.i.d samples as input to solve the second-stage problem. The solutions of 100 runs for each instance are summarised as a box plot and are presented in Figures 3a-3c. The average values of the objective function (total expected cost) for the 100 runs of each instance are also shown in Figure 3d.
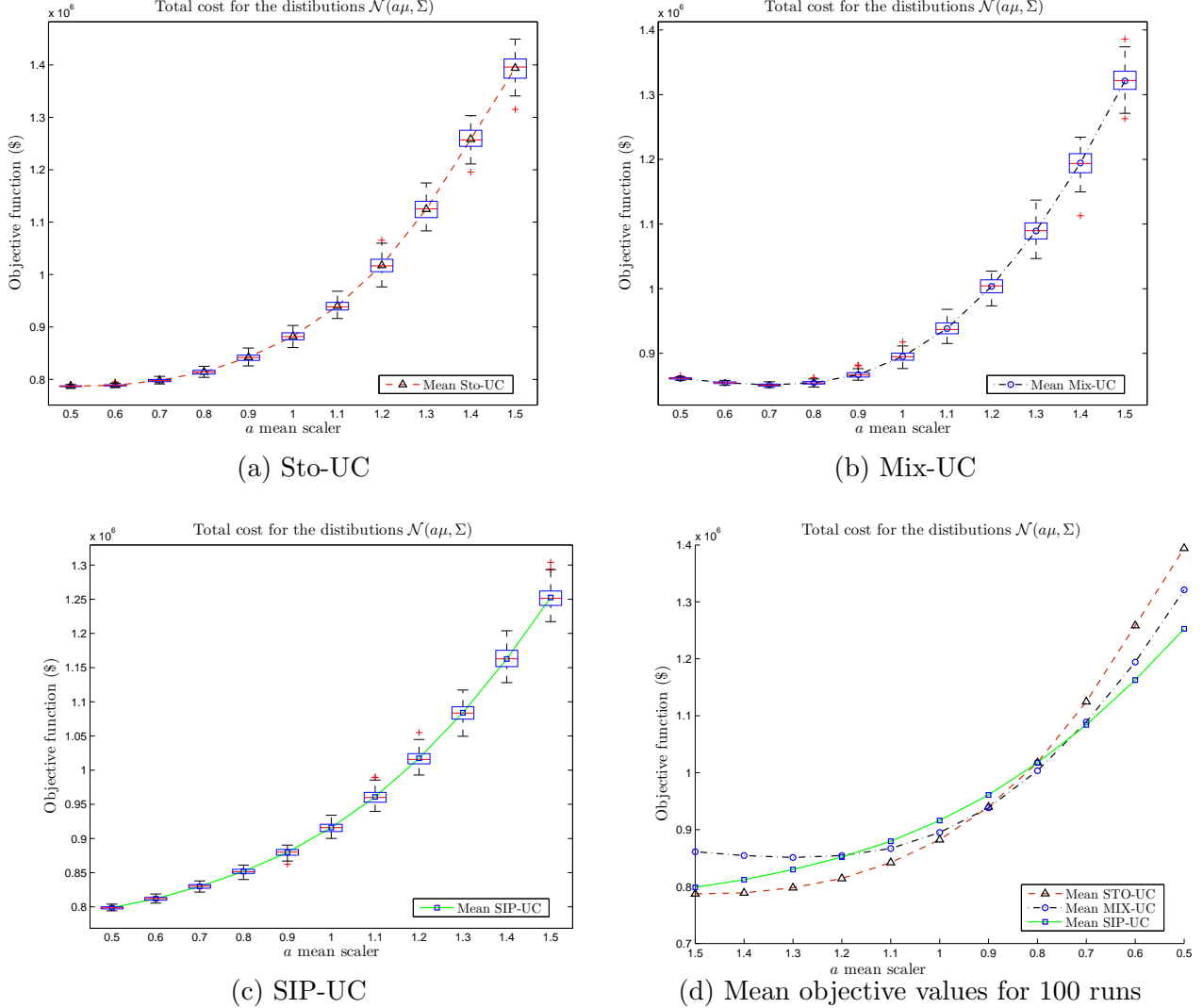

(a) Sto-UC

(b) Mix-UC

(c) SIP-UC

(d) Mean objective values for 100 runs

Figure 3: Sensitivity of solutions to variation in mean

It can be observed that when the actual wind distribution is $\mathcal{N} = (\boldsymbol{\mu}, \Sigma)$, i.e. when $a = 1$, the Sto-UC solution performs better than Mix-UC and SIP-UC solutions as expected since the assumed distribution in Sto-UC coincides with the actual distribution. The Sto-UC solution also performs better for all the instances with the mean of distribution greater than $\boldsymbol{\mu}$, i.e. when the mean of wind level is greater than expected in Sto-UC model. This is due to the ability of the system operator to dispatch higher levels of energy using the excess wind power rather than utilizing the costly reserve. On the other hand, the Mix-UC (SIP-UC) solution performs better when the mean of the wind distribution is less than 90% (80%) of

the anticipated value in the Sto-UC model. In other words the Mix-UC and SIP-UC robust solutions have lower total costs than the Sto-UC solution when the wind output is less than what decision maker assumed under the Sto-UC model.

In the second set of sensitivity test instances, we compare the performances of the Sto-UC, Mix-UC, and SIP-UC solutions to the possible changes in the covariances of the wind distribution.
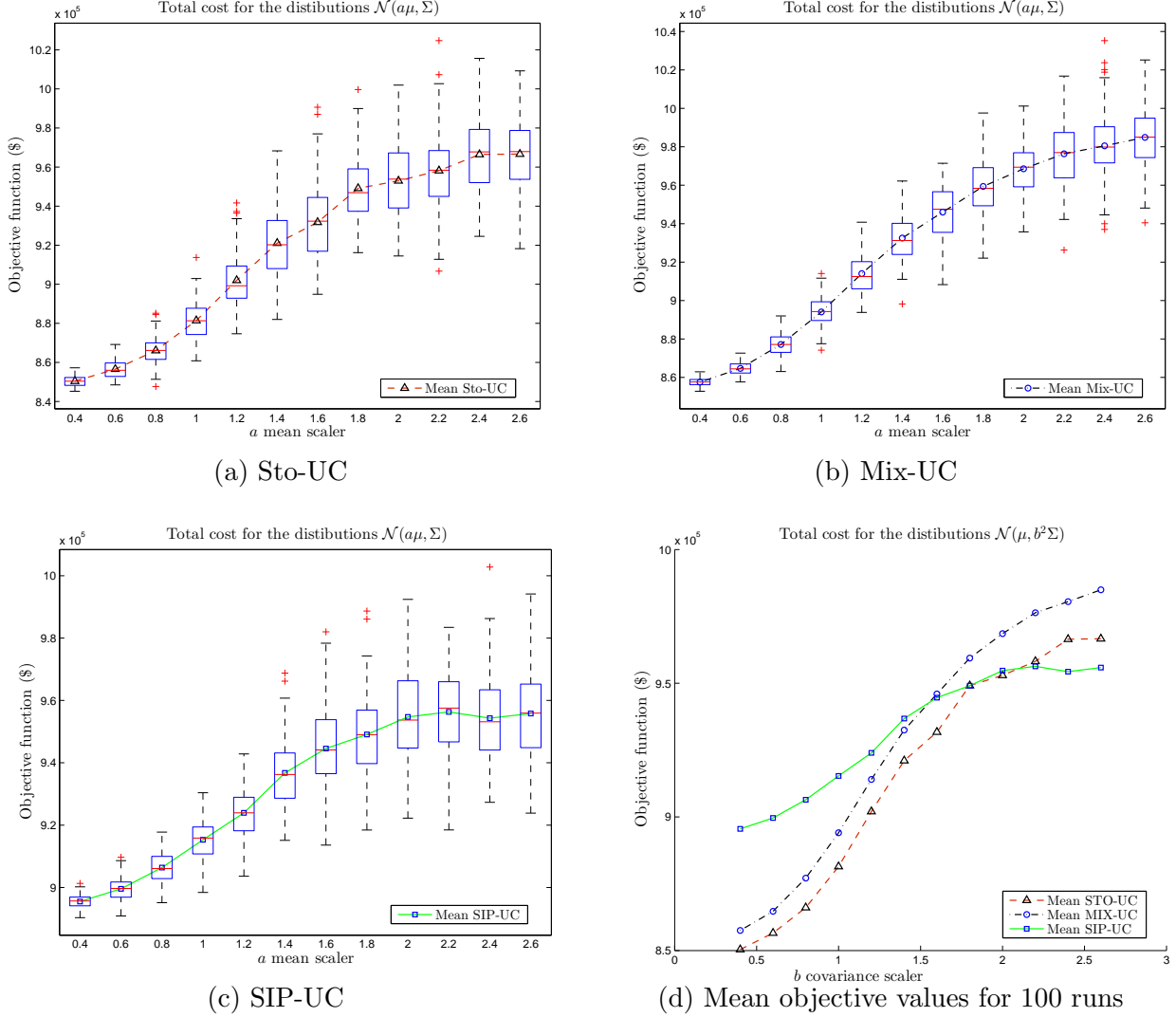

(a) Sto-UC

(b) Mix-UC

(c) SIP-UC

(d) Mean objective values for 100 runs

Figure 4: Sensitivity of solutions to variation in covariance

We specifically consider the following truncated normal distributions for the second-stage uncertain wind $\mathcal{N}(\boldsymbol{\mu}, (b)^2\Sigma)$, where $b = \{0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 2.2, 2.4, 2.6\}$. Similar to previous tests, we draw 150 i.i.d samples from the corresponding distribution in each instance and repeat each instance for 100 runs. The corresponding box plots for test instances for all three models are presented in Figures 4a-4c. Furthermore, the mean value of the objective function for 100 tests in each instance and each model is shown in Figure 4d. It can be observed that the SIP-UC solution has the least sensitivity to change in covariance of the distribution, since the only available information for the SIP-UC model was the first

moment condition of the distribution and there was no assumption on the covariance of the distribution. However, this additional flexibility comes at a cost and it can be observed that the SIP-UC solution is more conservative and costly for instances with covariance coefficients around 2, when compared to the Sto-UC solutions. For the instances with $b \geq 2.2$ the SIP-UC performs better than both the Sto-UC and Mix-UC solutions.

The sensitivity of the Mix-UC and Sto-UC solutions to changes in covariance are very similar and the difference between the two curves is almost unchanged across the test instances. This is due to the covariance assumptions in construction of the mixture model, i.e. covariance of all distributions used to construct the mixture model was equal to Σ. We have also constructed the mixture distribution using the following distributions:

$$\mathcal{N}\big(\mu, (0.8)^2 \Sigma\big), \mathcal{N}\big(\mu, \Sigma\big), \mathcal{N}\big(\mu, (1.2)^2 \Sigma\big).$$

The solution of this mixture model was very similar to the solutions of the Sto-UC model.

The final sensitivity test that we have performed is to randomly draw the wind input from historical Danish wind data. Specifically, we perform 100 independent runs, each includes 150 daily wind speed data drawn out of 263 historical daily wind speed by using sampling with replacement. For each of these 100 independent runs, the first stage solutions of the STO-UC, the MIX-UC and the SIP-UC are tested for performance against the corresponding 150 daily wind data. We then report the corresponding total costs of these solutions on the 100 runs. Figure 5a shows the box-plot of these total costs and Figure 5b shows their quantile plots. These quantile plots provide us with the relative comparison between STO-UC, MIX-UC and SIP-UC on different quantiles. For example, at quantile 1, the figure shows us the worst total cost among these three strategies while at quantile 0 it shows the best cases. We can see from the Figure 5b that MIX-UC does not performs well compared to the others. This could be because of the belief on mixed distribution does not apply to the Danish wind data. The SIP-UC, on the other hand, always outperforms STO-UC in all the quantiles. This implies that using only a belief on the mean is more 'robust' against sampling with replacement on the Danish wind data.
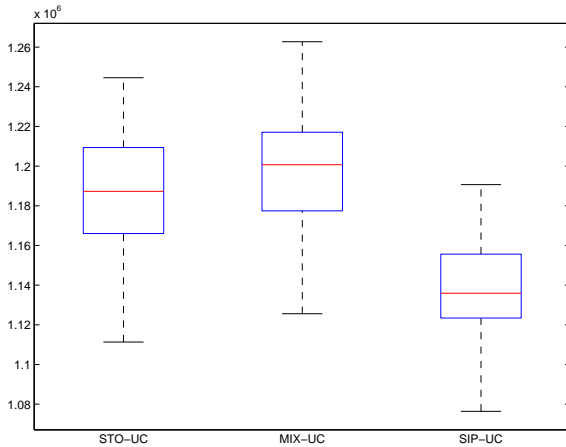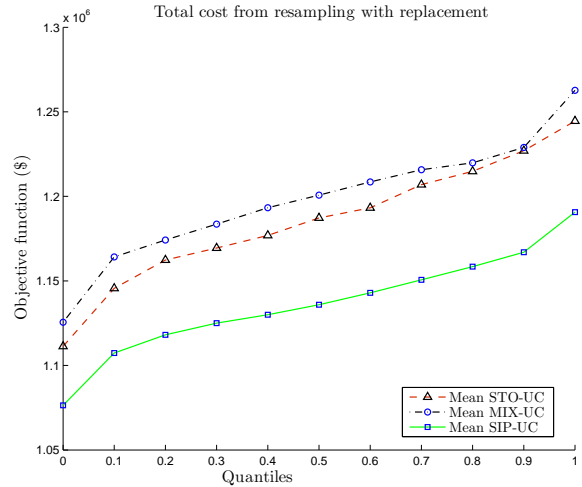
(a) Mean objective values for 100 runs   (b) Mean objective values for 100 runs

Figure 5: Sensitivity of solutions to variation in samples from historical wind data

Medium and Large Cases

We have also tested the limit of the models computationally by considering a medium-sized problem with 50 generator and a large-sized problem with 100 generators. The time period T is set to equal to 12. We set the stopping criteria as either 5 hours computer-running time or an optimality gap of no more than 1%, whichever reaches first. Sensitivity results for these cases are included in the Appendix B. Table 8 shows the computational times and optimality gaps on these two instances in addition to the small instance with 10 generators.

	STO-UC	MIX-UC	SIP-UC
$(n = 10, t = 24)$	539.16 seconds	196.7 seconds	1894 seconds
	0.79% optimality gap	0.91% optimality gap	0.5% optimality gap
$(n = 50, t = 12)$	2457.51 seconds	1050.46 seconds	18000 seconds
	0.94% optimality gap	0.84% optimality gap	2.09% optimality gap
$(n = 100, t = 12)$	18000 seconds	6202.39 seconds	$\cdots$
	1.01% optimality gap	0.99% optimality gap	$\cdots$

Table 8: Computational performance on different problem sizes.

We can see that the computational time increases significantly when the number of generator increases. This is because the number of binary variables, the number of continuous variables and the number of constraints in each of the STO-UC, MIX-UC and SIP-UC increase proportionally with the number of generators. For the case of 50 generators, the STO-UC can reach 0.94% optimality gap within around 40 minutes. The MIX-UC took less than 18 minutes to reach 0.85% optimality gap while the STO-UC stops after 5 hours when it reached 2.09% optimality gaps. For the case of 100 generators, only the MIX-UC reach the 1% optimality gap stopping criteria after 104 minutes while the STO-UC stops slightly higher at 1.01% optimality gap after 5 hours. There was no sign of the SIP-UC to reach reasonable optimality gap after 5 hours.

One interesting observation is that the MIX-UC actually took less time compared to the STO-UC. This is a bit counter-intuitive as the distributionally robust optimization models are build on the two-stage with one extra layer of optimization, i.e. a min-max-min problem with a worst-case max operator in the middle, and hence, we would generally expect it to be more difficult to solve than the stochastic counterpart. In fact, through dualization, the reformulated models for the MIX-UC and the SIP-UC are actually not much bigger than the stochastic counterpart. Specifically, the MIX-UC has one additional variables and two additional constraints compared to the STO-UC. The corresponding numbers of additional variables and constraints for the SIP-UC are 25 and 150, respectively.

We have programmed the entire models for the STO-UC, MIX-UC and SIP-UC into MAT-LAB and calling the CPLEX solver directly. It is possible to apply techniques for solving stochastic programming (e.g. by using Bender's decomposition) to improve the performance of these models and we expect that these could improve the numerical computation significantly. Using that method, we envisage that the increase in the number of generators would not impose significant computational issues as the number of binary variables only grow linearly. Nevertheless, we leave this enhanced methods for future research direction.

6   Conclusion

In this paper, we present a two-stage distributionally robust model that provides a novel and practical approach to deal with the uncertainty of the distribution of random wind output in the unit commitment problem. The model includes the technical ramping constraints as well as reliability condition against failure of up to one generating unit. The robustness takes into account the available information on uncertainty in two alternative ways. First, we assume that only the first-moment information of the random wind is given and use duality theory to formulate the problem as a linear semi-infinite program. The SIP model is then solved using sampling because the structure of the problem does not allow us to reformulate the semiinfinite constraints as semidefinite constraints. We establish exponential rate of convergence of the optimal value when the randomization scheme is applied to discretize the semi-infinite constraints.

Second, we assume that the information on probability distribution of uncertain wind is received through various sources, and we construct a mixture model to include these into decision making. The mixture model is also reformulated using duality theory and solved through the sample average approximation approach. Empirical tests have been carried out using an illustrative UC case study, taken from the literature, in order to illustrate the performance of the proposed robust models. The robust UC solutions may lose the potential of utilizing the wind power in high-wind climate; however, they perform much better in a low-wind climate as compared to the solutions that do not consider the uncertainty of the distribution (Sto-UC).

As one of the referees observed, the model has some limitations in absence of network constraints and second order moment information. In the unit commitment literature, transmission network has a very important impact on the UC decisions, particularly when there is significant wind power generation. Moreover, transmission line outages are an important factor for the (n-1) security constraints. Likewise, if we interpret the first order moment μ as a persistence forecast for each time period of the planning horizon, then the forecast routines used to generate μ may constitute standard deviations. We envisage that these are new directions for further research of the model but require significant new work on numerical schemes to cope with the additional complexity from network constraints and higher order moment conditions.

Acknowledgments

References

[1] Hassene Aissi, Cristina Bazgan, and Daniel Vanderpooten. Min–max and min–max regret versions of combinatorial optimization problems: A survey. *European Journal of Operational Research*, 197(2):427–438, 2009.

[2] Edward Anderson, Huifu Xu, and Dali Zhang. Confidence levels for cvar risk measures and minimax limits.

[3] José M Arroyo and Francisco D Galiana. Energy and reserve pricing in security and network-constrained electricity markets. *IEEE Transactions on Power Systems*, 20(2): 634–643, 2005.

[4] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88(3):411–424, 2000. ISSN 0025-5610.

[5] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton Univ Pr, 2009. ISBN 0691143684.

[6] D Bertsimas, X V Doan, K Natarajan and C P Teo, Models for Minimax Stochastic Linear Optimization Problems with Risk Aversion. Mathematics of Operations Research, 35, no. 3 (2010): 580-602

[7] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, pages 35–53, 2004. ISSN 0030-364X.

[8] Dimitris Bertsimas and Ioana Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804, 2005.

[9] Dimitris Bertsimas, Eugene Litvinov, Xu Andy Sun, Jinye Zhao, and Tongxin Zheng. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems*, 28(1):52–63, 2013. ISSN 0885-8950.

[10] François Bouffard and Francisco D Galiana. An electricity market with a probabilistic spinning reserve criterion. *IEEE Transactions on Power Systems*, 19(1):300–307, 2004.

[11] François Bouffard and Francisco D Galiana. Stochastic security for operations planning with significant wind power generation. In *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, pages 1–11. IEEE, 2008.

[12] François Bouffard, Francisco D Galiana, and José M Arroyo. Umbrella contingencies in security-constrained optimal power flow. In *15th Power Systems Computation Conference, PSCC*, volume 5, 2005.

[13] François Bouffard, Francisco D Galiana, and Antonio J Conejo. Market-clearing with stochastic security-part I: formulation. *IEEE Transactions on Power Systems*, 20(4): 1818–1826, 2005.

[14] François Bouffard, Francisco D Galiana, and Antonio J Conejo. Market-clearing with stochastic security-part II: Case studies. *IEEE Transactions on Power Systems*, 20(4): 1827–1835, 2005.

[15] Giuseppe Calafiore and Marco C Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[16] Pierre Carpentier, G Gohen, J-C Culioli, and Arnaud Renaud. Stochastic optimization of unit commitment: a new decomposition framework. *IEEE Transactions on Power Systems*, 11(2):1067–1073, 1996.

[17] Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3):595–612, 2010.

[18] Darinka Dentcheva and Werner Römisch. *Optimal power generation under uncertainty via stochastic programming*. Springer, 1998.

[19] Joel Goh and Melvyn Sim. Distributionally robust optimization and its tractable approximations. *Operations research*, 58(4-part-1):902–917, 2010.

[20] Ruiwei Jiang, Jianhui Wang, and Yongpei Guan. Robust unit commitment with wind power and pumped storage hydro. *IEEE Transactions on Power Systems*, 27(2):800–810, 2012.

[21] Efthymios Karangelos and François Bouffard. Towards full integration of demand-side resources in joint forward energy/reserve electricity markets. *IEEE Transactions on Power Systems*, 27(1):280–289, 2012.

[22] Panos Kouvelis and Gang Yu. *Robust discrete optimization and its applications*, volume 14. Springer, 1997.

[23] Henry J Landau. *Moments in mathematics*, volume 37. American Mathematical Soc., 1987.

[24] Yongchao Liu and Huifu Xu. Entropic approximation for mathematical programs with robust equilibrium constraints. *SIAM Journal on Optimization*, 2014.

[25] Yongchao Liu, Werner Römisch, and Huifu Xu. Quantitative stability analysis of stochastic generalized equations. *SIAM Journal on Optimization*, 2014.

[26] Sanjay Mehrotra and Dávid Papp. A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization. *SIAM Journal on Optimization*, 24(4):1670–1697, 2014.

[27] Arkadi Nemirovski. Lectures on robust convex optimization. *class notes, Georgia Inst. of Technol., Fall*, 2009.

[28] František Nožička. *Theorie der linearen parametrischen Optimierung*, volume 24. Akademie-Verlag, 1974.

[29] Narayana Prasad Padhy. Unit commitment – a bibliographical survey. *IEEE Transactions on Power Systems*, 19(2):1196–1205, 2004.

[30] Anthony Papavasiliou, Shmuel S Oren, and Richard P O'Neill. Reserve requirements for wind power integration: A scenario-based stochastic programming framework. *IEEE Transactions on Power Systems*, 26(4):2197–2206, 2011. ISSN 0885-8950.

[31] David Peel and Geoffrey J McLachlan. Robust mixture modelling using the t distribution. *Statistics and computing*, 10(4):339–348, 2000.

[32] David Pozo and Javier Contreras. A chance-constrained unit commitment with an $n - K$ security criterion and significant wind generation. *IEEE Transactions on Power Systems*, 28(3):2842–2851, 2013.

[33] Qianfan Wang, Yongpei Guan, and Jianhui Wang. A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output. *IEEE Transactions on Power Systems*, 27(1):206–215, 2012.

[34] Jose F Restrepo and Francisco D Galiana. Assessing the yearly impact of wind power through a new hybrid deterministic/stochastic unit commitment. *IEEE Transactions on Power Systems*, 26(1):401–410, 2011.

[35] Tyrrell R. Rockafellar. *Conjugate Duality and Optimization*, volume 16. SIAM, 1974.

[36] Tyrrell R. Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[37] Herbert Scarf, KJ Arrow, and S Karlin. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production*, 10:201–209, 1958.

[38] Alexander Shapiro. On duality theory of conic linear problems. In *Semi-infinite programming*, pages 135–165. Springer, 2001.

[39] A. Shapiro, D. Dentcheva and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, MPS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics, SIAM, Philadelphia, PA, 2009.

[40] Alexander Shapiro and Huifu Xu. Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. *Optimization*, 57(3):395–418, 2008.

[41] A. L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157, 1973.

[42] A. Street, F. Oliveira, and J. M. Arroyo. Contingency-constrained unit commitment with $n - K$ security criterion: A robust optimization approach. *IEEE Transactions on Power Systems*, 26(3):1581–1590, 2011. ISSN 0885-8950. doi: Doi10.1109/Tpwrs.2010.2087367.

[43] Samer Takriti, John R Birge, and Erik Long. A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems*, 11(3):1497–1508, 1996.

[44] A. Tuohy, P. Meibom, E. Denny, and M. O'Malley. Unit commitment for systems with significant wind penetration. *IEEE Transactions on Power Systems*, 24(2):592–601, 2009. ISSN 0885-8950.

[45] David W Walkup, Roger J-B Wets, et al. Lifting projections of convex polyhedra. *Pacific Journal of Mathematics*, 28(2):465–475, 1969.

[46] J. Wang, M. Shahidehpour, and Z. Li. Security-constrained unit commitment with volatile wind power generation. *IEEE Transactions on Power Systems*, 23(3):1319–1327, 2008. ISSN 0885-8950.

[47] Wolfram Wiesemann, Daniel Kuhn, and Berç Rustem. Robust resource allocations in temporal networks. *Mathematical programming*, 135(1-2):437–471, 2012.

[48] Wolfram Wiesemann, Daniel Kuhn, and Berç Rustem. Robust markov decision processes. *Mathematics of Operations Research*, 38(1):153–183, 2013.

[49] David Wozabal. Robustifying convex risk measures for linear portfolios: A nonparametric approach. *Operations Research*, 62(6):1302–1315, 2014.

[50] Huifu Xu and Dali Zhang. Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications. *Mathematical Programming*, 119 (2):371–401, 2009.

[51] Shushang Zhu and Masao Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations research*, 57(5):1155–1168, 2009.

A Ramp constraints and reliability and $n-1$ security constraint

Ramp constraints

For conventional generation units, it is important to take into account the short-term dynamics of generation output over the consecutive periods. Such requirements are often referred to as ramp constraints which limit the maximum increase or decrease of generated power from one time period to the next, reflecting the thermal and mechanical inertia which need to be overtaken in order for the generating unit to increase or decrease its output.

In a two-stage stochastic framework, we need to define the ramp constraints for any possible realization of the net load ξ. At the second stage after the realization of the uncertain net load, the scheduled energy q_{it} does not change and the generating units adapt their production to accommodate the realized net load. To do this, up and down reserves $(\hat{r}_{it}^{up}(\xi), \hat{r}_{it}^{dw}(\xi))$ are used. We denote the actual power output of unit i at time t and scenario ξ by $\hat{q}_{it}(\xi)$ which is defined as follows:

$$\hat{q}_{it}(\xi) = q_{it} + \hat{r}_{it}^{up}(\xi) - \hat{r}_{it}^{dw}(\xi), \qquad \forall t, i, \xi.$$

Note that up and down reserves should not be simultaneously positive. We consider the ramp constraints for all four combinations of on-off decisions between any two consecutive periods $t-1$ and t as follows,

1. If $u_{i(t-1)} = 1$ and $u_{it} = 1$, then the generating unit i is coupled for both hours, and ramp limitations in hour t are bounded by the ramp up and down rates, RU_i and RD_i, respectively. The ramp constraints will then be as follows:

$$\hat{q}_{it}(\xi) - \hat{q}_{i(t-1)}(\xi) \leq RU_i, \qquad\qquad \forall t, i, \xi, \qquad\qquad \text{(R1)}$$
$$\hat{q}_{i(t-1)}(\xi) - \hat{q}_{it}(\xi) \leq RD_i, \qquad\qquad \forall t, i, \xi. \qquad\qquad \text{(R2)}$$

2. If $u_{i(t-1)} = 0$ and $u_{it} = 1$, generating unit i starts up at the beginning of period t and hence the limitation for hour t should be the starting up ramp SU_i. The ramp constraints will then be

$$\hat{q}_{it}(\xi) - \hat{q}_{i(t-1)}(\xi) \leq SU_i, \qquad\qquad \forall t, i, \xi, \qquad\qquad \text{(R1)}$$
$$\hat{q}_{i(t-1)}(\xi) - \hat{q}_{it}(\xi) \leq RD_i, \qquad\qquad \forall t, i, \xi. \qquad\qquad \text{(R2)}$$

Note that (R2) always holds because $\hat{q}_{i(t-1)}(\xi) = 0$, $\hat{q}_{it}(\xi) \geq 0$ and $RD_i \geq 0$.

3. If $u_{i(t-1)} = 1$ and $u_{it} = 0$, generating unit i shutdowns at the beginning of time period t and hence limitation during the hour t should be the shut-down ramp SD_i. The ramp constraints then become

$$\hat{q}_{it}(\xi) - \hat{q}_{i(t-1)}(\xi) \leq RU_i, \qquad\qquad \forall t, i, \xi, \qquad\qquad \text{(R1)}$$
$$\hat{q}_{i(t-1)}(\xi) - \hat{q}_{it}(\xi) \leq SD_i, \qquad\qquad \forall t, i, \xi. \qquad\qquad \text{(R2)}$$

Note that (R1) always holds because the left hand side is always negative whereas the right hand side is always positive.

4. If $u_{i(t-1)} = 0$ and $u_{it} = 0$, then generating unit i is off during both periods and the ramp constraints are

$$\hat{q}_{it}(\xi) - \hat{q}_{i(t-1)}(\xi) \leq SU_i, \qquad\qquad \forall t, i, \xi, \qquad\qquad \text{(R1)}$$
$$\hat{q}_{i(t-1)}(\xi) - \hat{q}_{it}(\xi) \leq SD_i, \qquad\qquad \forall t, i, \xi. \qquad\qquad \text{(R2)}$$

A summary of ramp limitations for all of the scenarios discussed above is presented in Table 9:

Table 9: Ramp limitations

	$u_{i(t-1)}$	u_t	RHS of (R1)	RHS of (R2)
Uncoupled	0	0	SU_i	SD_i
Starting up	0	1	SU_i	RD_i
Shutting down	1	0	RU_i	SD_i
Coupled	1	1	RU_i	RD_i

Having defined the ramp constraints for all on/off decision scenarios over two time periods, we can generalize them in relation to decision variable u_{it} as follow,

$$\hat{q}_{it}(\xi) - \hat{q}_{i(t-1)}(\xi) \leq RU_i u_{i(t-1)} + SU_i(1 - u_{i(t-1)}), \qquad \forall t, i, \xi,$$

$$\hat{q}_{i(t-1)}(\xi) - \hat{q}_{it}(\xi) \leq RD_i u_{it} + SD_i(1 - u_{it}), \qquad \forall t, i, \xi.$$

Reliability and $n - 1$ security constraint

To ensure the secure operation of the system under the failure of up to one scheduled generator, we consider the worst-case scenario that the system could possibly face; that is, for any given time period, the generator with the highest total scheduled generation and actual up reserve fails under the lowest level of wind (highest net load) available. This can be modeled as

$$\sum_i (q_{it} + r_{it}^{up}) - \max_i \{q_{it} + \hat{r}_{it}^{up}(\xi)\} + S_t(\xi) \geq d_t - \min_\xi \{w_t(\xi)\}, \qquad \forall t, \xi. \qquad (A.1)$$

In the literature of energy management, this is known as $n - 1$ criteria. Let us denote the worst (lowest) realization of wind output at time t by

$$\underline{\xi}_t = \arg \min_\xi \{w_t(\xi)\},$$

and let $S_t(\underline{\xi})$ denote the corresponding load shedding for this realization. Since the actual up reserve used has to be less than the scheduled up reserve, i.e. $\hat{r}_{it}^{up}(\xi) \leq r_{it}^{up}$ for every scenario ξ, the inequality (A.1) can be rewritten as

$$\sum_i (q_{it} + r_{it}^{up}) - \max_i \{q_{it} + r_{it}^{up}\} + S_t(\underline{\xi}) \geq d_t - w_t(\underline{\xi}), \qquad \forall t. \qquad (A.2)$$

Let us denote the upper limit of total generation and reserve schedules at time t as

$$Q_t = \sum_i (q_{it} + r_{it}^{up}), \qquad \forall t.$$

The inequality constraint (A.2) can be rewritten as

$$Q_t - (q_{it} + r_{it}^{up}) + S_t(\underline{\xi}) \geq l_t(\underline{\xi}), \qquad \forall t, i.$$

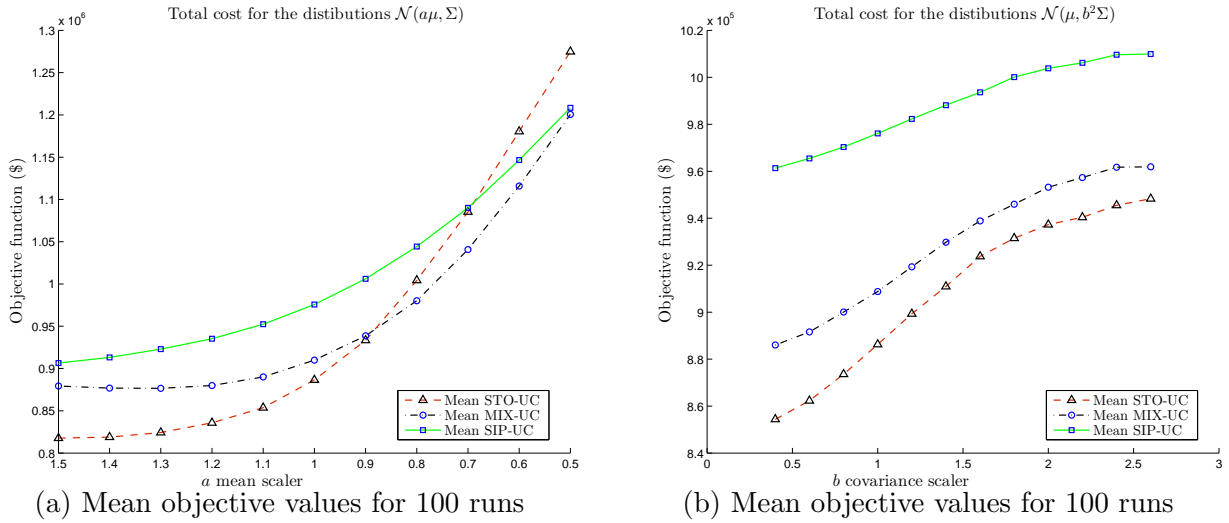# B Numerical Results for Medium and Large Cases

(a) Mean objective values for 100 runs
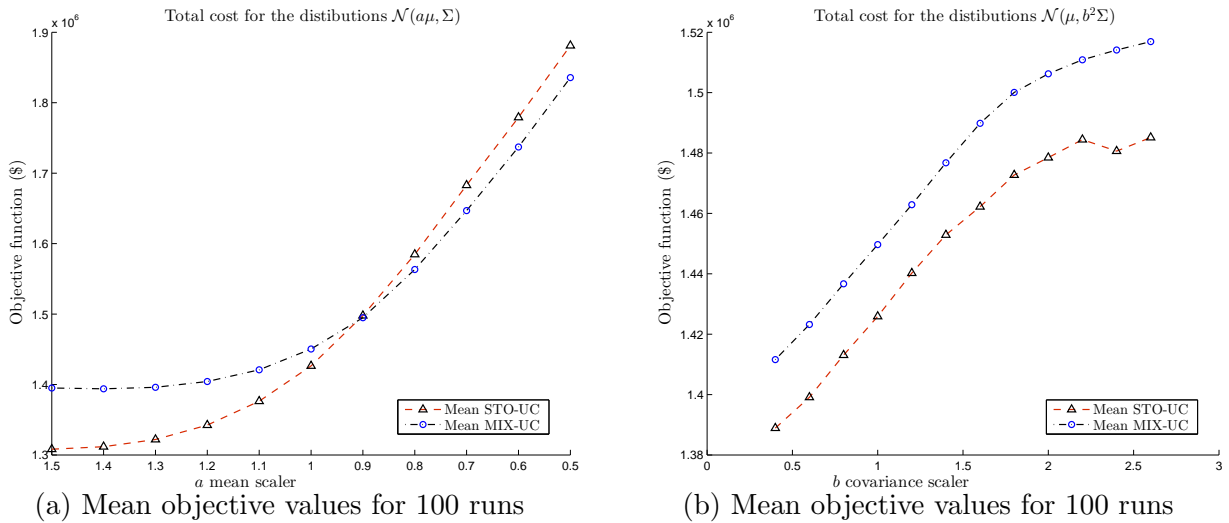
(b) Mean objective values for 100 runs

Figure 6: Sensitivity of solutions to variation in mean and variance for a medium-sized problem with 50 generators (with $T = 12$, $S = 150$.)

(a) Mean objective values for 100 runs

(b) Mean objective values for 100 runs

Figure 7: Sensitivity of solutions to variation in mean and variance for a medium-sized problem with 100 generators (with $T = 12$, $S = 150$.)