A Two-Stage Moment Robust Optimization Model and its Solution Using Decomposition*

Sanjay Mehrotra† and He Zhang‡

July 23, 2013

Abstract

Moment robust optimization models formulate a stochastic problem with an uncertain probability distribution of parameters described by its moments. In this paper we study a two-stage stochastic convex programming model using moments to define the probability ambiguity set for the objective function coefficients in both stages. A decomposition based algorithm is given. We show that this two-stage model can be solved to any precision in polynomial time. We consider a special case where the probability ambiguity sets are described by the exact information of the first two moments and the convex functions are piece-wise linear utility functions. A two-stage stochastic semidefinite programming formulation is given of this problem and we provide computational results on the performance of this problem using a portfolio optimization application. Results show that the two stage modeling is effective when forecasting models have predictive power.

1 Introduction

Moment robust optimization models specify information on the distribution of the uncertain parameters using moments of the probability distribution of these parameters. The probability distribution is not known. Scarf [16] proposed such a model for the newsvendor problem. In his problem the given information is the mean and variance of the distribution. Recently, different forms of the distributional ambiguity sets have been considered. Bertsimas et al. [2] studied a piece-wise linear utility model with exact knowledge of the first two moments. Two cases are considered in [2]: (i) the uncertain coefficients are in the objective function, and (ii) the uncertain coefficients are in the right-hand side of the constraints. For the first case, Bertsimas et al. [2] give an equivalent semidefinite programming (SDP) formulation. When the uncertain coefficients are in the right-hand side of the constraints, they show that their robust model is NP-complete. The uncertain parameters only appear in the second stage problem, hence their model can be considered as a single stage moment robust optimization

*Research partially supported by NSF grant CMMI-1100868 and ONR grant N00014-09-10518 and N00014210051.

†Dept. of Industrial Engineering & Management Sciences, Northwestern University, Evanston, IL 60201

‡Dept. of Industrial Engineering & Management Sciences, Northwestern University, Evanston, IL 60201

problem. Delage and Ye [5] also considered such a single stage moment robust convex optimization program with the ambiguity set defined by a confidence region for both first and second moments. They show that this problem can be formulated as a semi-infinite convex optimization problem. It is then solved by the ellipsoid method in polynomial time. Alternatives to using moments to specify the distribution uncertainty have been proposed. Pflug and Wozabal [14] analyzed the portfolio selection problem with the ambiguity set defined by a confidence region over a reference probability measure using the Kantorovich distance. Shapiro and Ahmed [17] analyzed a class of convex optimization problem with ambiguity set defined by general moment constraints and bounds over the probability measures. Mehrotra and Zhang [13] give conic reformulations of ambiguity models in [5, 14, 17] for the distributionally robust least-squares problem.

In this paper we consider a two-stage moment robust stochastic convex optimization problem given as:

$$\min_{\mathbf{x}\in\mathcal{X}} \; f(\mathbf{x}) + G(\mathbf{x}), \tag{1.1}$$

$$G(\mathbf{x}) = \sum_{k=1}^{K} \pi_k G_k(\mathbf{x}), \tag{1.2}$$

$$G_k(\mathbf{x}) := \min_{\mathbf{w}_k\in\mathcal{W}_k(\mathbf{x})} \; g_k(\mathbf{w}_k). \tag{1.3}$$

The objective functions $f(\mathbf{x})$ and $g_k(\mathbf{w}_k)$ are defined as:

$$f(\mathbf{x}) := \max_{\mathbb{P}\in\mathscr{P}_1} \mathbb{E}_{\mathbb{P}}[\rho_1(\mathbf{x}, \tilde{\mathbf{p}})], \tag{1.4}$$

$$g_k(\mathbf{w}_k) := \max_{\mathbb{P}\in\mathscr{P}_{2,k}} \mathbb{E}_{\mathbb{P}}[\rho_2(\mathbf{w}_k, \tilde{\mathbf{q}})], \tag{1.5}$$

where $\rho_1(\cdot)$ and $\rho_2(\cdot)$ are two general functions and the expectations in (1.4) and (1.5) are taken for the random vectors $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$. $\mathscr{S}_1$ and $\mathscr{S}_2$ are first and second stage sample spaces. Note that $\mathscr{P}_{2,k}$ may depend on k. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be the measures defined on $\mathscr{S}_1$ and $\mathscr{S}_2$ with the Borel σ-algebra. The probability ambiguity sets $\mathscr{P}_1$ and $\mathscr{P}_{2,k}$ are defined as:

$$\begin{aligned} \mathscr{P}_1 := \{\mathbb{P} : \mathbb{P} \in \mathcal{M}_1, \mathbb{E}_{\mathbb{P}}[\mathbf{1}] = \mathbf{1}, (\mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{p}}] - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1} (\mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{p}}] - \boldsymbol{\mu}_1) \le \alpha_1, \\ \mathbb{E}_{\mathbb{P}}[(\tilde{\mathbf{p}} - \boldsymbol{\mu}_1)(\tilde{\mathbf{p}} - \boldsymbol{\mu}_1)^T] \preceq \beta_1 \boldsymbol{\Sigma}_1\}, \end{aligned} \tag{1.6}$$

$$\begin{aligned} \mathscr{P}_{2,k} := \{\mathbb{P} : \mathbb{P} \in \mathcal{M}_2, \mathbb{E}_{\mathbb{P}}[\mathbf{1}] = \mathbf{1}, (\mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{q}}] - \boldsymbol{\mu}_{2,k})^T \boldsymbol{\Sigma}_{2,k}^{-1} (\mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{q}}] - \boldsymbol{\mu}_{2,k}) \le \alpha_{2,k}, \\ \mathbb{E}_{\mathbb{P}}[(\tilde{\mathbf{q}} - \boldsymbol{\mu}_{2,k})(\tilde{\mathbf{q}} - \boldsymbol{\mu}_{2,k})^T] \preceq \beta_{2,k} \boldsymbol{\Sigma}_{2,k}\}, \end{aligned} \tag{1.7}$$

where $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{\mu}_{2,k}$, $\boldsymbol{\Sigma}_{2,k}$ are moment parameters used to define ambiguity sets $\mathscr{P}_1$ and $\mathscr{P}_{2,k}$. This definition is same as the one used in Delage and Ye [5]. The feasible set $\mathcal{X}$ is a convex compact set and the second stage feasible set $\mathcal{W}_k(\mathbf{x})$ is defined as:

$$\mathcal{W}_k(\mathbf{x}) := \{\mathbf{W}_k \mathbf{w}_k = \mathbf{h}_k - \mathbf{T}_k \mathbf{x}, \mathbf{w}_k \in \mathcal{W}\}, \tag{1.8}$$

where $\mathcal{W}$ is a nonempty convex compact set in $\mathbb{R}^{n_2}$, $\mathbf{x} \in \mathbb{R}^{n_1}$ and $\mathbf{w}_k \in \mathbb{R}^{n_2}$. The situation leading to modeling framework (1.1)-(1.8) is one where the current parameters of a decision problem are uncertain, and this uncertainty propogates through a known stochastic process, leading to future uncertain parameters. For example, the estimated current returns of stocks are ambiguous, and the portfolio needs to be balanced at a future step where the returns are also ambiguous. In (1.1)-(1.3), we have moment robust problems in both stages and the parameter estimations of the moments for the second stage problem (1.3) are uncertain when the decision makers optimize the first-stage problem (1.1). Note that although the parameter $\mathbf{W}_k$, $\mathbf{T}_k$ and $\mathbf{h}_k$ are random in this model, in view of the NP-completeness results for the single stage case [2], in the current paper we do not define an ambiguity set for these parameters. Only the parameters of the convex objective function for the first and the second stage problem are defined over an ambiguity set.

In this paper we develop a decomposition based algorithm to solve (1.1)-(1.8). We show that (1.1)-(1.8) can be solved in polynomial time under suitable, yet general assumptions. For the single stage moment robust problem in [5], a key requirement is that a semi-infinite constraint is verified, or an oracle gives a cut in polynomial time. Since the constraint can only be verified to ϵ-precision in our case, we need further development to prove the polynomial solvability of (1.1)-(1.8). The ϵ-precision only allows an ϵ-precision verification of the constraint feasibility. Also, we can only generate an approximate cut to separate infeasible points. Both facts suggest that we need to use approximate separation oracles to prove the polynomial solvability. We also study a two-stage moment robust portfolio optimization model and empirical results suggests that the two-stage modeling framework is effective when we have forecasting power.

This paper is organized as follows. In Section 2 we give additional notations, the necessary definitions, and assumptions for (1.1)-(1.8). In Section 3 we develop an equivalent formulation of the two-stage framework (1.1)-(1.3). In Section 4, we present some results for convex optimization problem needed to calculate ϵ-sub and supergradients. Section 5 gives analysis of an ellipsoidal decomposition algorithm for (1.1)-(1.8). In Section 6 a two-stage moment robust portfolio optimization model is considered. We use data to study the effectiveness of the two-stage model. We compare our two-stage model with two other models and experimental results suggest that our two-stage model has better performance when we have forecasting power.

2 Definitions and Assumptions

In this section we summarize several definitions and assumptions used in the rest of this paper. Throughout we use the phrase "polynomial time" to refer to computations performed using number of elementary operations that are polynomial in problem dimension, size of the input data, and $\log(\frac{1}{\epsilon})$, using exact arithmetic. Here ϵ is the desired precision for the optimization problem (1.1)-(1.8).

Definition 1 *Let us consider variables* $\mathbf{x}$ *with* $\dim(\mathbf{x}) = n$. *For any set* $C \subseteq \mathbf{R}^n$ *and a positive real number* ϵ, *the set* $B_{\mathbf{x}}(C, \epsilon)$ *is defined as:*

$$B_{\mathbf{x}}(C, \epsilon) := \{\mathbf{x} \in \mathbf{R}^n : ||\mathbf{x} - \mathbf{y}|| \leq \epsilon \text{ for some } \mathbf{y} \in C\}.$$

$B_{\mathbf{x}}(C, \epsilon)$ *is the* ϵ*-ball covering of* C. *In particular, when* C *is a singleton, i.e.* $C = \{\bar{\mathbf{x}}\}$, *we set* $B_{\mathbf{x}}(C, \epsilon) = B_{\mathbf{x}}(\bar{\mathbf{x}}, \epsilon)$, *which is the* ϵ*-ball around* $\bar{\mathbf{x}}$.

Definition 2 *Let us consider an optimization problem* (P)*:* $\min_{\mathbf{x}\in C} h(\mathbf{x})$*, where* $C \subseteq \mathbf{R}^n$ *is a full dimensional closed convex set and* $h(\mathbf{x})$ *is a convex function of* $\mathbf{x}$*. We say that* (P) *is solved to* ϵ*-precision if we can find a feasible* $\bar{\mathbf{x}} \in C$*, such that* $h(\bar{\mathbf{x}}) \leq h(\mathbf{x}) + \epsilon$ *for all* $\mathbf{x} \in C$*.*

Definition 3 *For a convex function* $f(\mathbf{x})$ *defined on* $\mathbb{R}^n$*,* $\mathbf{d} \in \mathbb{R}^n$ *is an* ϵ*-subgradient at* $\mathbf{x}$ *if for all* $\mathbf{z} \in \mathbb{R}^n$*,* $f(\mathbf{z}) \geq f(\mathbf{x}) + \mathbf{d}^T(\mathbf{z} - \mathbf{x}) - \epsilon$*. For a concave function* $f(\mathbf{x})$ *defined on* $\mathbb{R}^n$*,* $\mathbf{d} \in \mathbb{R}^n$ *is an* ϵ*-supergradient at* $\mathbf{x}$ *if for all* $\mathbf{z} \in \mathbb{R}^n$*,* $f(\mathbf{z}) \leq f(\mathbf{x}) + \mathbf{d}^T(\mathbf{z} - \mathbf{x}) + \epsilon$*.*

Definition 4 *Let us consider the convex optimization problem*

$$\begin{aligned} \min\ & f(\mathbf{x}) \\ \text{s.t. } & \mathbf{g}(\mathbf{x}) \leq 0 \\ & \mathbf{h}(\mathbf{x}) = 0 \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{2.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$*,* $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ *are convex component-wise,* $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^l$ *is affine, and* $\mathcal{X}$ *is a nonempty convex compact set. For any* $\boldsymbol{\mu} \in \mathbb{R}^m_+$*,* $\boldsymbol{\lambda} \in \mathbb{R}^l$*, we define the Lagrangian dual as:*

$$\theta(\boldsymbol{\mu}, \boldsymbol{\lambda}) := \min_{\mathbf{x}\in\mathcal{X}} \{f(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x})\}. \tag{2.2}$$

Let γ^* *be the optimal objective value of* (2.1)*. We call* $\bar{\boldsymbol{\mu}} \in \mathbb{R}^m_+$ *and* $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^l$ *to be an* ϵ *optimal Lagrange multipliers of the constraints* $\mathbf{g}(\mathbf{x}) \leq 0$ *and* $\mathbf{h}(\mathbf{x}) = 0$ *if* $0 \leq \gamma^* - \theta(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) < \epsilon$*.*

Note that because of weak duality $\gamma^* - \theta(\boldsymbol{\mu}, \boldsymbol{\lambda}) \geq 0$ for any $\boldsymbol{\mu} \in \mathbb{R}^m_+$ and $\boldsymbol{\lambda} \in \mathbb{R}^l$. For the two-stage moment robust problem (1.1)-(1.3), we make the following assumptions:

Assumption 1 *For* $\alpha_1 \geq 0$*,* $\beta_1 \geq 1$*, and* $\boldsymbol{\Sigma}_1 \succ 0$*,* $\rho_1(\mathbf{x}, \tilde{\mathbf{p}})$ *is* $\mathbb{P}$*-integrable for all* $\mathbb{P} \in \mathscr{P}_1$*.*

Assumption 2 *The sample space* $\mathscr{S}_1 \subset \mathbb{R}^{m_1}$ *is convex and compact (closed and bounded), and it is equipped with an oracle that for any* $\mathbf{p} \in \mathbb{R}^{m_1}$ *can either confirm that* $\mathbf{p} \in \mathscr{S}_1$ *or provide a hyperplane that separates* $\mathbf{p}$ *from* $\mathscr{S}_1$ *in polynomial time.*

Assumption 3 *The set* $\mathcal{X} \subset \mathbb{R}^{n_1}$ *is convex, compact, and full dimensional (closed and bounded with nonempty interior). There exists an* $\mathbf{x}_0 \in \mathcal{X}$ *and* $r_0^1, R_0^1 \in \mathbb{R}_+$*, such that* $B_{\mathbf{x}}(\mathbf{x}_0, r_0^1) \subseteq \mathcal{X} \subseteq B_{\mathbf{x}}(0, R_0^1)$*. In the context of the ellipsoid method we assume that* $\mathbf{x}_0$*,* r_0^1 *and* R_0^1 *are known.* $\mathcal{X}$ *is equipped with an oracle that for any* $\mathbf{x} \in \mathbb{R}^{n_1}$ *can either confirm that* $\bar{\mathbf{x}} \in \mathcal{X}$ *or provide a vector* $\mathbf{d} \in \mathbb{R}^{n_1}$ *with* $||\mathbf{d}||_\infty \geq 1$ *such that* $\mathbf{d}^T\bar{\mathbf{x}} < \mathbf{d}^T\mathbf{x}$ *for* $\forall \mathbf{x} \in \mathcal{X}$ *in polynomial time.*

Assumption 4 *The function* $\rho_1(\mathbf{x}, \mathbf{p})$ *is concave in* $\mathbf{p}$*. In addition, given a pair* $(\mathbf{x}, \mathbf{p})$*, it is assumed that in polynomial time, one can:*

1. *evaluate the value of* $\rho_1(\mathbf{x}, \mathbf{p})$*;*
2. *find a supergradient of* $\rho_1(\mathbf{x}, \mathbf{p})$ *in* $\mathbf{p}$*.*

Assumption 5 *The function* $\rho_1(\mathbf{x}, \mathbf{p})$ *is convex in* $\mathbf{x}$*. In addition, it is assumed that one can find in polynomial time a subgradient of* $\rho_1(\mathbf{x}, \mathbf{p})$ *in* $\mathbf{x}$*.*

Assumption 6 *For any* $k \in \{1, \ldots, K\}$, $\alpha_{2,k} \geq 0$, $\beta_{2,k} \geq 1$, *and* $\mathbf{\Sigma}_{2,k} \succ 0$, $\rho_2(\mathbf{w}, \mathbf{q})$ *is* $\mathbb{P}$*-integrable for all* $\mathbb{P} \in \mathscr{P}_{2,k}$.

Assumption 7 *The sample space* $\mathscr{S}_2 \subset \mathbb{R}^{m_2}$ *is convex and compact (closed and bounded), and it is equipped with an oracle that for any* $\mathbf{q} \in \mathbb{R}^{m_2}$ *can either confirm that* $\mathbf{q} \in \mathscr{S}_2$ *or provide a hyperplane that separates* $\mathbf{q}$ *from* $\mathscr{S}_2$ *in polynomial time.*

Assumption 8 *The set* $\mathcal{W} \subset \mathbb{R}^{n_2}$ *is convex, compact, and full dimensional. There exists known* $\mathbf{w}_0 \in \mathcal{W}$ *and* $r_0^2, R_0^2 \in \mathbb{R}_+$, *such that* $B_{\mathbf{w}}(\mathbf{w}_0, r_0^2) \subseteq \mathcal{W} \subseteq B_{\mathbf{w}}(0, R_0^2)$. *It is equipped with an oracle that for any* $\mathbf{w} \in \mathbb{R}^{n_2}$ *can either confirm that* $\mathbf{w} \in \mathcal{W}$ *or provide a hyperplane, i.e. a vector* $\mathbf{d} \in \mathbb{R}^{n_2}$ *with* $||\mathbf{d}||_\infty \geq 1$ *that separates* $\mathbf{w}$ *from* $\mathcal{W}$ *in polynomial time.*

Assumption 9 *For any* $k \in \{1, \ldots, K\}$, $\mathbf{x} \in \mathcal{X}$, $\mathcal{W}_k(\mathbf{x}) := \{\mathbf{w}_k : \mathbf{W}_k \mathbf{w}_k = \mathbf{h}_k - \mathbf{T}_k \mathbf{x}, \mathbf{w}_k \in \mathcal{W}\} \neq \emptyset$.

Assumption 9 implies that $\mathcal{W}_k(\mathbf{x})$ is nonempty and compact for any $k \in \{1, \ldots, K\}$. This is a standard assumption in stochastic programming. It may be ensured by using an artificial variable for each scenario.

Assumption 10 *The function* $\rho_2(\mathbf{w}, \mathbf{q})$ *is concave in* $\mathbf{q}$. *In addition, given a pair* $(\mathbf{w}, \mathbf{q})$, *we assume that we can in polynomial time:*

1. *evaluate the value of* $\rho_2(\mathbf{w}, \mathbf{q})$;
2. *find a supergradient of* $\rho_2(\mathbf{w}, \mathbf{q})$ *in* $\mathbf{q}$.

Assumption 11 *The function* $\rho_2(\mathbf{w}, \mathbf{q})$ *is convex in* $\mathbf{w}$. *In addition, we assume that we can find in polynomial time a subgradient of* $\rho_2(\mathbf{w}, \mathbf{q})$ *in* $\mathbf{w}$.

Assumptions 1-2 are to guarantee that the first stage ambiguity set (1.6) is well defined. Assumption 3 requires that the first stage feasible region is compact, and a separating hyperplane with enough norm can be generated for any infeasible point. Assumptions 4-5 make sure that the objective function is convex/conave and its sub and supergradients can be computed efficiently. Assumptions 1-5 are similar as the assumptions in [5]. Assumptions 6-11 are similar to Assumptions 1-5. These assumptions ensure that the second stage problem is a well-defined convex program for each scenario.

3 Equivalent Formulation of Two-Stage Moment Robust Problem

In this section we give an equivalent formulation of the two stage moment robust problem (1.1)-(1.3). The next theorem gives such a reformulation for a single stage problem.

Theorem 1 (Delage and Ye 2010 [5]). *Let* $\mathbf{v} := (\mathbf{Y}, \mathbf{y}, y_0, t)$ *and define:*

$$c(\mathbf{v}) := y_0 + t \tag{3.1}$$

$$\begin{aligned} \mathcal{V}(h, \mathbf{x}) := \{\mathbf{v} :\ & y_0 \geq h(\mathbf{x}, \mathbf{q}) - \mathbf{q}^T \mathbf{Y} \mathbf{q} - \mathbf{q}^T \mathbf{y},\ \forall \mathbf{q} \in \mathscr{S}, \\ & t \geq (\beta \mathbf{\Sigma}_0 + \boldsymbol{\mu_0} \boldsymbol{\mu_0}^T) \bullet \mathbf{Y} + \boldsymbol{\mu}_0^T \mathbf{y} + \sqrt{\alpha} \left|\left| \mathbf{\Sigma}_0^{\frac{1}{2}} (\mathbf{y} + 2\mathbf{Y} \boldsymbol{\mu}_0) \right|\right|, \\ & \mathbf{Y} \succeq 0\}. \end{aligned} \tag{3.2}$$

Consider the moment robust problem

$$\min_{\mathbf{x}\in\mathcal{X}}(\max_{\mathbb{P}\in\mathscr{P}} \mathbb{E}_{\mathbb{P}}[h(\mathbf{x},\tilde{\mathbf{q}})]) \tag{3.3}$$

with probability ambiguity set $\mathscr{P}$ *defined as*

$$\mathscr{P} := \{\mathbb{P} : \mathbb{P} \in \mathcal{M}, \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{q}}] = 1, (\mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{q}}] - \boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{q}}] - \boldsymbol{\mu}_0) \leq \alpha, \\ \mathbb{E}_{\mathbb{P}}[(\tilde{\mathbf{q}} - \boldsymbol{\mu}_0)(\tilde{\mathbf{q}} - \boldsymbol{\mu}_0)^T] \preceq \beta\boldsymbol{\Sigma}_0\}. \tag{3.4}$$

Consider the following assumptions:

(i) $\alpha \geq 0$, $\beta \geq 1$, $\boldsymbol{\Sigma} \succ 0$, *and* $h(\mathbf{x},\mathbf{q})$ *is* $\mathbb{P}$*-integrable for all* $\mathbb{P} \in \mathscr{P}$.

(ii) *The sample space* $\mathscr{S} \subset \mathbb{R}^m$ *is convex and compact, and it is equipped with an oracle that for any* $\mathbf{q} \in \mathbb{R}^m$ *can either confirm that* $\mathbf{q} \in \mathscr{S}$ *or provide a hyperplane that separates* $\mathbf{q}$ *from* $\mathscr{S}$ *in polynomial time.*

(iii) *The feasible region* $\mathcal{X}$ *is convex and compact, and it is equippend with an oracle that for any* $\mathbf{x} \in \mathbb{R}^n$ *can either confirm that* $\mathbf{x} \in \mathcal{X}$ *or provide a hyperplane that separates* $\mathbf{x}$ *from* $\mathcal{X}$ *in polynomial time.*

(iv) *The function* $h(\mathbf{x},\mathbf{q})$ *is concave in* $\mathbf{q}$. *In addition, given a pair* $(\mathbf{x},\mathbf{q})$, *it is assumed that one in polynomial time can:*
1. *evaluate the value of* $h(\mathbf{x},\mathbf{q})$;
2. *find a supergradient of* $h(\mathbf{x},\mathbf{q})$ *in* $\mathbf{q}$.

(v) *The function* $h(\mathbf{x},\mathbf{q})$ *is convex in* $\mathbf{x}$. *In addition, it is assumed that one can find in polynomial time a subgradient of* $h(\mathbf{x},\mathbf{q})$ *in* $\mathbf{x}$.

If assumption (i) *is satisfied, then for any given* $\mathbf{x} \in \mathcal{X}$, *the optimal value of the inner problem* $\max_{\mathbb{P}\in\mathscr{P}} \mathbb{E}_{\mathbb{P}}[h(\mathbf{x},\tilde{\boldsymbol{q}})]$ *in* (3.3) *is equal to the optimal value* $c(\mathbf{v}^*)$ *of the problem:*

$$\min_{\mathbf{v}\in\mathcal{V}(h,\mathbf{x})} c(\mathbf{v}). \tag{3.5}$$

If assumptions (i)-(v) *are satisfied, then* (3.3) *is equivalent to the following problem*

$$\min_{\mathbf{v}\in\mathcal{V}(h,\mathbf{x}),\mathbf{x}\in\mathcal{X}} c(\mathbf{v}). \tag{3.6}$$

Problem (3.6) *is well defined, and can be solved by the ellipsoid method to any precision* ϵ *in polynomial time.* ∎

Applying Theorem 1 to (1.4) and (1.5), we have an equivalent two-stage semi-infinite programming formulation of (1.1)-(1.8) as stated in the following theorem.

Theorem 2 *Suppose that Assumptions* 1-11 *are satisfied. Then the two-stage moment robust problem* (1.1)-(1.8) *is equivalent to*

$$\min_{\mathbf{x}\in\mathcal{X},\mathbf{v}^1\in\mathcal{V}(\rho_1,\mathbf{x})} c(\mathbf{v}^1) + G(\mathbf{x}) \tag{3.7}$$

$$G(\mathbf{x}) = \sum_{k=1}^{K} \pi_k G_k(\mathbf{x}), \tag{3.8}$$

$$G_k(\mathbf{x}) := \min_{\mathbf{w}_k\in\mathcal{W}_k(\mathbf{x}),\mathbf{v}_{2,k}\in\mathcal{V}(\rho_2,\mathbf{w}_k)} c(\mathbf{v}_{2,k}), \tag{3.9}$$

where:

$$\mathbf{v}_1 := (\mathbf{Y}_1, \mathbf{y}_1, y_1^0, t_1), \tag{3.10}$$
$$c(\mathbf{v}_1) := y_1^0 + t_1, \tag{3.11}$$
$$\begin{aligned}\mathcal{V}_1(\rho_1, \mathbf{x}) := \{\mathbf{v}_1 :\ & y_1^0 \geq \rho_1(\mathbf{x}, \mathbf{p}) - \mathbf{p}^T \mathbf{Y}_1 \mathbf{p} - \mathbf{p}^T \mathbf{y}_1,\ \forall \mathbf{p} \in \mathscr{S}_1 \\ & t_1 \geq (\beta_1 \mathbf{\Sigma}_1 + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^T) \bullet \mathbf{Y}_1 + \boldsymbol{\mu}_1^T \mathbf{y}_1 + \sqrt{\alpha_1} \left|\left| \mathbf{\Sigma}_1^{\frac{1}{2}} (\mathbf{y}_1 + 2\mathbf{Y}_1 \boldsymbol{\mu}_1) \right|\right| \\ & \mathbf{Y}_1 \succeq 0\},\end{aligned} \tag{3.12}$$

and for any $k \in \{1, \ldots, K\}$,

$$\mathbf{v}_{2,k} := (\mathbf{Y}_{2,k}, \mathbf{y}_{2,k}, y_{2,k}^0, t_{2,k}), \tag{3.13}$$
$$c(\mathbf{v}_{2,k}) := y_{2,k}^0 + t_{2,k}, \tag{3.14}$$
$$\begin{aligned}\mathcal{V}_{2,k}(\rho_2, \mathbf{w}_k) := \{\mathbf{v}_{2,k} :\ & y_{2,k}^0 \geq \rho_2(\mathbf{w}_k, \mathbf{q}) - \mathbf{q}^T \mathbf{Y}_{2,k} \mathbf{q} - \mathbf{q}^T \mathbf{y}_{2,k},\ \forall \mathbf{q} \in \mathscr{S}_2, \\ & t_{2,k} \geq (\beta_2 \mathbf{\Sigma}_{2,k} + \boldsymbol{\mu}_{2,k} \boldsymbol{\mu}_{2,k}^T) \bullet \mathbf{Y}_{2,k} + \boldsymbol{\mu}_{2,k}^T \mathbf{y}_{2,k} \\ & + \sqrt{\alpha_2} \left|\left| \mathbf{\Sigma}_{2,k}^{\frac{1}{2}} (\mathbf{y}_{2,k} + 2\mathbf{Y}_{2,k} \boldsymbol{\mu}_{2,k}) \right|\right|, \mathbf{Y}_{2,k} \succeq 0\}.\end{aligned} \tag{3.15}$$

Problem (3.7)-(3.9) is a two-stage stochastic program, where both stages are semi-infinite programming problems. We will develop a decomposition algorithm to prove the polynomial solvability of (3.7)-(3.9) in Section 5.

4 Subgradient Calculation and Polynomial Solvability

In this section, we summarize some known results on convex optimization. We also present a basic result for computing an ϵ-subgradient for general convex optimization problem. The following theorem from Grotschel, Lovasz and Schrijver [7] shows that a convex optimization problem and the separation problem of a convex set are polynomially equivalent.

Theorem 3 (Grotschel et al. [7, Theorem 3.1]). *Consider a convex optimization problem of the form*

$$\min_{\mathbf{z} \in \mathscr{Z}} \mathbf{c}^T \mathbf{z}$$

with a linear objective function and a convex closed feasible set $\mathscr{Z} \subset \mathbb{R}^n$. *Assume that there are known constants* $\mathbf{a}_0, r$ *and* R *such that* $B_{\mathbf{z}}(\mathbf{a}_0, r) \subseteq \mathscr{Z} \subseteq B_{\mathbf{z}}(0, R)$. *Assume that we have an oracle such that given a vector* $\bar{\mathbf{z}}$ *and a number* $\delta > 0$, *we can conclude with one of the following:*

1. $\bar{\mathbf{z}}$ *passes the test of the oracle, i.e., ensure that* $\bar{\mathbf{z}} \in B_{\mathbf{z}}(\mathscr{Z}, \delta)$ *in polyhnomial time.*

2. $\bar{\mathbf{z}}$ *does not pass the test of the oracle and it can generate a vector* $\mathbf{d} \in \mathbf{R}^n$ *with* $||d||_\infty \geq 1$ *such that* $\mathbf{d}^T \bar{\mathbf{z}} \leq \mathbf{d}^T \mathbf{z} + \delta$ *for every* $\mathbf{z} \in \mathscr{Z}$ *in polynomial time.*

Then, given an $\epsilon > 0$, *we can find a vector* $\mathbf{y}$ *satisfying the oracle with some* $\delta \leq \epsilon$ *such that* $\mathbf{c}^T \mathbf{y} - \epsilon \leq \mathbf{c}^T \mathbf{z}$ *for* $\forall \mathbf{z} \in \mathscr{Z}$ *in polynomial time by using the ellipsoid method.*

The following proposition tells us that the average of N ϵ-subgradients is still an ϵ-subgradient of the average of the N convex functions.

Proposition 1 (Hiriart-Urruty and Lemarechal [8, Theorem 3.1.1]). *If $h_1(\mathbf{x}), \ldots, h_N(\mathbf{x})$ are N convex functions w.r.t. $\mathbf{x}$ and $\nabla h_1^\epsilon(\bar{\mathbf{x}}), \ldots, \nabla h_N^\epsilon(\bar{\mathbf{x}})$ are ϵ-subgraduents of these N convex function at $\bar{\mathbf{x}}$, then for any $\pi_1, \ldots, \pi_N \in \mathbb{R}_+$ with $\sum_{i=1}^N \pi_i = 1$, $\sum_{i=1}^N \pi_i \nabla h_i^\epsilon(\bar{\mathbf{x}})$ is a ϵ-subgradient of $\sum_{i=1}^N \pi_i h_i(\mathbf{x})$ at $\bar{\mathbf{x}}$.* ∎

The following lemma shows that the ϵ-optimal solutions of the Lagrangian may be used to obtain an ϵ-supergradient of the Lagrangian w.r.t. the Lagrange multipliers.

Lemma 1 *Consider the convex programming problem defined in (2.1), where $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ is convex, $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^l$ is affine, and $\mathcal{X}$ is a nonempty convex compact set. Assume that $\{\mathbf{x} : \mathbf{g}(\mathbf{x}) < 0\} \cap \{\mathbf{x} : \mathbf{h}(\mathbf{x}) = 0\} \cap \mathcal{X} \neq \varnothing$. For any given $\bar{\boldsymbol{\mu}} \in \mathbb{R}_+^m$, $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^l$ and $\epsilon > 0$, if $\bar{\mathbf{x}}$ is an ϵ-optimal solution of $\theta(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}})$, i.e,*

$$-\epsilon < \theta(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) - (f(\bar{\boldsymbol{x}}) + \bar{\boldsymbol{\mu}}^T \mathbf{g}(\bar{\mathbf{x}}) + \bar{\boldsymbol{\lambda}}^T \mathbf{h}(\bar{\mathbf{x}}) < 0,$$

then $(\mathbf{g}(\bar{\mathbf{x}}); \mathbf{h}(\bar{\mathbf{x}}))$ is an ϵ-supergradient of $\theta(\cdot, \cdot)$ at $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}})$, where $\theta(\cdot, \cdot)$ is the Lagrangian dual defined in (2.2).

Proof It is well known that $\theta(\boldsymbol{\mu}, \boldsymbol{\lambda})$ is a concave function w.r.t. $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ [3, Sec. 5.1.2]. Since f, $\mathbf{g}$ are convex, $\mathbf{h}$ is affine and $\mathcal{X}$ is nonempty and compact, we know that $\theta(\boldsymbol{\mu}, \boldsymbol{\lambda})$ is finite for $\forall (\boldsymbol{\mu}; \boldsymbol{\lambda}) \in \mathbb{R}^{m+l}$. Therefore,

$$\begin{aligned}
\theta(\boldsymbol{\mu}, \boldsymbol{\lambda}) &= \min_{\mathbf{x} \in \mathcal{X}} \{f(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x})\} \\
&\leq f(\bar{\mathbf{x}}) + \boldsymbol{\mu}^T \mathbf{g}(\bar{\mathbf{x}}) + \boldsymbol{\lambda}^T \mathbf{h}(\bar{\mathbf{x}}) \\
&= f(\bar{\mathbf{x}}) + [(\boldsymbol{\mu}; \boldsymbol{\lambda}) - (\bar{\boldsymbol{\mu}}; \bar{\boldsymbol{\lambda}})]^T (\mathbf{g}(\bar{\mathbf{x}}); \mathbf{h}(\bar{\mathbf{x}})) + (\bar{\boldsymbol{\mu}}; \bar{\boldsymbol{\lambda}})^T (\mathbf{g}(\bar{\mathbf{x}}); \mathbf{h}(\bar{\mathbf{x}})) \\
&\leq \theta(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) + [(\boldsymbol{\mu}; \boldsymbol{\lambda}) - (\bar{\boldsymbol{\mu}}; \bar{\boldsymbol{\lambda}})]^T (\mathbf{g}(\bar{\mathbf{x}}); \mathbf{h}(\bar{\mathbf{x}})) + \epsilon \qquad \blacksquare
\end{aligned}$$

The following strong duality theorem is from Bazaraa, Sherali and Shetty [1, Theorem 6.2.4].

Theorem 4 *Consider the convex optimization problem (2.1). Let $\mathcal{X}$ be a nonempty convex set in $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ be convex, and let $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^l$ be affine; that is, $\mathbf{h}$ is of the form $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$. Suppose that the Slater constraint qualification holds, i.e, there exists an $\hat{\mathbf{x}} \in \mathcal{X}$ such that $\mathbf{g}(\hat{\mathbf{x}}) < 0$, $\mathbf{h}(\hat{\mathbf{x}}) = 0$, and $\mathbf{0} \in int(\mathbf{h}(\mathcal{X}))$, where $\mathbf{h}(\mathcal{X}) = \{\mathbf{h}(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ and $int(\cdot)$ is the interior of a set. Then,*

$$\inf\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \leq 0, \mathbf{h}(\mathbf{x}) = 0, \mathbf{x} \in \mathcal{X}\} = \sup\{\theta(\boldsymbol{\mu}, \boldsymbol{\lambda}) : \boldsymbol{\mu} \geq 0\}.$$

Furthermore, if the inf *is finite, then* $\sup\{\theta(\boldsymbol{\mu}, \boldsymbol{\lambda}) : \boldsymbol{\mu} \geq 0\}$ *is achieved at $(\bar{\boldsymbol{\mu}}; \bar{\boldsymbol{\lambda}})$ with $\bar{\boldsymbol{\mu}} \geq 0$. If the* inf *is achieved at $\bar{\mathbf{x}}$, then $\bar{\boldsymbol{\mu}}^T \mathbf{g}(\bar{\mathbf{x}}) = 0$.* ∎

The following theorem states that the ϵ-optimal Lagrangian multipliers can be used as an ϵ-subgradient of the perturbed convex optimization problem (4.1) at the origin.

Theorem 5 *Consider the convex optimization problem (2.1). Let $\mathcal{X}$ be a nonempty convex set in $\mathbb{R}^n$, let $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ be convex, and $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^l$ be affine. Assume that the Slater constaint qualification as in Theorem 4 holds. Assume that $\bar{\boldsymbol{\mu}} \in \mathbb{R}_+^m$, $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^l$ are*

ϵ-optimal Lagrangian multipliers for the constraints $\mathbf{g}(\mathbf{x}) \leq 0$ and $\mathbf{h}(\mathbf{x}) = 0$. Now consider the perturbed problem:

$$\begin{aligned} \pi(\mathbf{u}, \mathbf{v}) := \min & f(\mathbf{x}) \\ \text{s.t. } & \mathbf{g}(\mathbf{x}) \leq \mathbf{u} \\ & \mathbf{h}(\mathbf{x}) = \mathbf{v} \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{4.1}$$

with $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^l$. Then, $\pi(\mathbf{u}, \mathbf{v})$ is convex and $(\bar{\boldsymbol{\mu}}; \bar{\boldsymbol{\lambda}})$ is an ϵ-subgradient of $\pi(\mathbf{u}, \mathbf{v})$ at $(\mathbf{0}, \mathbf{0})$ w.r.t. $(\mathbf{u}; \mathbf{v})$.

Proof The convexity of $\pi(\mathbf{u}, \mathbf{v})$ is known from [3, Sec. 5.6.1]. For any $\boldsymbol{\mu} \in \mathbb{R}^m_+$, $\boldsymbol{\lambda} \in \mathbb{R}^l$, the Lagrangian function of the original problem $\pi(\mathbf{0}, \mathbf{0})$ is written as

$$L(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}).$$

Slater constraint qualification conditions ensure that the strong duality holds. For given $\boldsymbol{\mu} \in \mathbb{R}^m_+$, $\boldsymbol{\lambda} \in \mathbb{R}^l$, consider the dual problem: $\theta(\boldsymbol{\mu}, \boldsymbol{\lambda}) := \inf_{\mathbf{x} \in \mathcal{X}} L(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})$. For any $\mathbf{u} \in \mathbb{R}^m$, $\mathbf{v} \in \mathbb{R}^l$, let

$$\mathcal{Y}(\mathbf{u}, \mathbf{v}) := \{\mathbf{x} : \mathbf{x} \in \mathcal{X}, \mathbf{g}(\mathbf{x}) \leq \mathbf{u}, \mathbf{h}(\mathbf{x}) = \mathbf{v}\}.$$

If $\mathcal{Y}(\mathbf{u}, \mathbf{v}) \neq \varnothing$, since $\theta(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \inf_{\mathbf{x} \in \mathcal{X}} \{f(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x})\}$, for any $\bar{\mathbf{x}} \in \mathcal{Y}(\mathbf{u}, \mathbf{v})$, we can have:

$$\begin{aligned} \theta(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) &= \inf_{\mathbf{x} \in \mathcal{X}} \{f(\mathbf{x}) + \bar{\boldsymbol{\mu}}^T \mathbf{g}(\mathbf{x}) + \bar{\boldsymbol{\lambda}}^T \mathbf{h}(\mathbf{x})\} \\ &\leq f(\bar{\mathbf{x}}) + \bar{\boldsymbol{\mu}}^T \mathbf{g}(\bar{\mathbf{x}}) + \bar{\boldsymbol{\lambda}}^T \mathbf{h}(\bar{\mathbf{x}}) \\ &\leq f(\bar{\mathbf{x}}) + \bar{\boldsymbol{\mu}}^T \mathbf{u} + \bar{\boldsymbol{\lambda}}^T \mathbf{v}. \end{aligned}$$

The first inequality follows from the definition of inf and the second inequality follows from $\bar{\boldsymbol{\mu}} \in \mathbb{R}^m_+$. Since $\bar{\mathbf{x}}$ is an arbitrary point in $\mathcal{Y}(\mathbf{u}, \mathbf{v})$, we can take the infimum of the right-hand side over the set $\mathcal{Y}(\mathbf{u}, \mathbf{v})$ to get:

$$\theta(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) \leq \pi(\mathbf{u}, \mathbf{v}) + \bar{\boldsymbol{\mu}}^T \mathbf{u} + \bar{\boldsymbol{\lambda}}^T \mathbf{v}.$$

Since $0 < \pi(\mathbf{0}, \mathbf{0}) - \theta(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) < \epsilon$, we know that:

$$\pi(\mathbf{0}, \mathbf{0}) - \epsilon \leq \theta(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\lambda}}) \leq \pi(\mathbf{u}, \mathbf{v}) + \bar{\boldsymbol{\mu}}^T \mathbf{u} + \bar{\boldsymbol{\lambda}}^T \mathbf{v}. \tag{4.2}$$

Inequality (4.2) holds when $\mathcal{Y}(\mathbf{u}, \mathbf{v}) = \varnothing$ because $\pi(\mathbf{u}, \mathbf{v}) = \infty$ in this case. Therefore, we can conclude that $(\bar{\boldsymbol{\mu}}; \bar{\boldsymbol{\lambda}})$ is an ϵ-subgradient of $\pi(\mathbf{u}, \mathbf{v})$ at $(\mathbf{0}, \mathbf{0})$ with respect to $(\mathbf{u}; \mathbf{v})$. ∎

5 A Decomposition Algorithm for a General Two-Stage Moment Robust Optimization Model

In Section 3 we presented an equivalent formulation (3.7)-(3.9) of the two-stage moment robust program (1.1)-(1.3) as a semi-infinite program. In this section we propose a decomposition algorithm to show that this equivalent formulation can be solved to any precision in polynomial time. A key ingredient of the decomposition algorithm is the construction of an ϵ-subgradient of the function $G_k(\mathbf{x})$ defined in (3.9). Theorem 5 will be useful for this purpose. First, observe that we can apply Theorem 1 to the second stage problem (3.9) to solve them in polynomial time. This is stated in the following corollary.

Corollary 1 *For* $\forall k \in \{1, \ldots, K\}$*, let Assumptions* 6-10 *be satisfied. The second stage problem* $G_k(\mathbf{x})$ *defined as* (3.9) *can be solved to any precision* ϵ *in polynomial time.*

Proof Given $k \in \{1, \ldots, K\}$, consider the moment robust problem:

$$\min_{\mathbf{w}_k \in \mathcal{W}_k(\mathbf{x})} \max_{\mathbb{P} \in \mathscr{P}_{2,k}} \mathbb{E}[\rho_2(\mathbf{w}_k, \tilde{\mathbf{q}})].$$

Assumption 8 guarantees that for $k \in \{1, \ldots, K\}$ and $\mathbf{x} \in \mathcal{X}$, the set $\mathcal{W}_k(\mathbf{x})$ is nonempty and bounded. This verifies condition (iii) in Theorem 1. Assumption 6, 7, 9, 10 verify conditions (i), (ii), (iv), (v). ∎

For any given $k \in \{1, \ldots, K\}$, define the Lagrangian function $\mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x})$ of (3.9) as:

$$\mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x}) := \min_{\mathbf{w}_k, \mathbf{Y}_{2,k}, \mathbf{y}_{2,k}, y^0_{2,k}, t_{2,k}} \quad y^0_{2,k} + t_{2,k} + \boldsymbol{\lambda}^T(\mathbf{W}_k \mathbf{w}_k - \mathbf{h}_k + \mathbf{T}_k \mathbf{x}) \tag{5.1a}$$

$$\text{s.t. } y^0_{2,k} \geq \rho_2(\mathbf{w}_k, \mathbf{q}) - \mathbf{q}^T \mathbf{Y}_{2,k} \mathbf{q} - \mathbf{q}^T \mathbf{y}_{2,k}, \ \forall \mathbf{q} \in \mathscr{S}_2 \tag{5.1b}$$

$$t_{2,k} \geq (\beta_{2,k} \boldsymbol{\Sigma}_{2,k} + \boldsymbol{\mu}_{2,k} \boldsymbol{\mu}_{2,k}^T) \bullet \mathbf{Y}_{2,k} + \boldsymbol{\mu}_{2,k}^T \mathbf{y}_{2,k} + \sqrt{\alpha_{2,k}} \left\| \boldsymbol{\Sigma}_{2,k}^{\frac{1}{2}} (\mathbf{y}_{2,k} + 2\mathbf{Y}_{2,k} \boldsymbol{\mu}_{2,k}) \right\| \tag{5.1c}$$

$$\mathbf{Y}_{2,k} \succeq 0, \ \mathbf{w}_k \in \mathcal{W}. \tag{5.1d}$$

We now give a summary of the remaining analysis in this section. In Proposition 2, we show that for any given $\boldsymbol{\lambda}$, the Lagrangian problem $\mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x})$ can be evaluated to any precision polynomially. In Proposition 3, the ϵ approximate solution of $\mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x})$ can be used to generate an ϵ-subgradient to prove the polynomial time solvability of the second stage problem for each given first stage solution $\mathbf{x}$, as stated in Lemma 2 and 3. Based on this result, we further prove the polynomial solvability of the two-stage semi-infinite problem (3.7)-(3.9) in Theorem 6.

The following proposition states that (5.1) can be solved to any precision in polynomial time.

Proposition 2 *Let Assumptions* 6-10 *be satisfied. Then, for* $\forall \boldsymbol{\lambda} \in \mathbb{R}^{l_2}$*,* $\mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x})$ *can be evaluated and a solution* $(\mathbf{w}_k, \mathbf{v}_{2,k})$ *can be found to any precision* ϵ *in polynomial time.*

Proof Given $k \in \{1, \ldots, K\}$, $\mathbf{x} \in \mathcal{X}$, and $\boldsymbol{\lambda} \in \mathbb{R}^{l_2}$, consider the optimization problem:

$$\min_{\mathbf{w}_k \in \mathcal{W}} \max_{\mathbb{P} \in \mathscr{P}_{2,k}} \mathbb{E}_{\mathbb{P}}[\phi_k(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{w}_k, \mathbf{q})], \tag{5.2}$$

where $\phi_k(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{w}_k, \mathbf{q}) := \rho_2(\mathbf{w}_k, \mathbf{q}) + \boldsymbol{\lambda}^T(\mathbf{W}_k \mathbf{w}_k - \mathbf{h}_k + \mathbf{T}_k \mathbf{x})$. Since $\boldsymbol{\lambda}^T(\mathbf{W}_k \mathbf{w}_k - \mathbf{h}_k + \mathbf{T}_k \mathbf{x})$ is linear w.r.t. $\mathbf{w}_k$, Assumptions 10 and 11 are also satisfied for $\phi_k(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{w}^{\boldsymbol{\xi}}, \mathbf{q})$. Therefore, by using Theorem 1, (5.2) is equivalent to

$$\begin{aligned} \min_{\mathbf{w}_k, \mathbf{Y}_{2,k}, \mathbf{y}_{2,k}, y^0_{2,k}, t_{2,k}} \quad & y^0_{2,k} + t_{2,k} \\ \text{s.t. } & y^0_{2,k} \geq \phi_k(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{w}_k, \mathbf{q}) - \mathbf{q}^T \mathbf{Y}_{2,k} \mathbf{q} - \mathbf{q}^T \mathbf{y}_{2,k}, \ \forall \mathbf{q} \in \mathscr{S}_2 \\ & t_{2,k} \geq (\beta_{2,k} \boldsymbol{\Sigma}_{2,k} + \boldsymbol{\mu}_{2,k} \boldsymbol{\mu}_{2,k}^T) \bullet \mathbf{Y}_{2,k} + \boldsymbol{\mu}_{2,k}^T \mathbf{y}_{2,k} \\ & \qquad + \sqrt{\alpha_{2,k}} \left\| \boldsymbol{\Sigma}_{2,k}^{\frac{1}{2}} (\mathbf{y}_{2,k} + 2\mathbf{Y}_{2,k} \boldsymbol{\mu}_{2,k}) \right\| \\ & \mathbf{Y}_{2,k} \succeq 0, \ \mathbf{w}_k \in \mathcal{W}, \end{aligned} \tag{5.3}$$

and (5.3) is polynomially solvable. By substituting for $\mathbf{y}_{2,k}$, $t_{2,k}$ in the objective, it is easy to see that (5.3) is equivalent to (5.1), which implies that (5.1) can be solved to any precision ϵ in polynomial time. ∎

The following proposition shows that the ϵ-optimal solution of the Lagrangian problem (5.1) gives an ϵ-supergradient of $\mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x})$ w.r.t. $\boldsymbol{\lambda}$.

Proposition 3 *For any $k \in \{1, \ldots, K\}$, $\mathbf{x} \in \mathcal{X}$, $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^{l_2}$, let $(\bar{\mathbf{w}}_k, \bar{\mathbf{v}}_{2,k}) := (\bar{\mathbf{w}}_k, \bar{\mathbf{Y}}_{2,k}, \bar{\mathbf{y}}_{2,k}, \bar{y}^0_{2,k}, \bar{t}_{2,k})$ be an ϵ-optimal solution of the Lagrangian problem defined in (5.1) for $\boldsymbol{\lambda} = \bar{\boldsymbol{\lambda}}$. Then, $\mathbf{W}_k \bar{\mathbf{w}}_k - \mathbf{h}_k - \mathbf{T}_k \mathbf{x}$ is an ϵ-supergradient of $\mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x})$ w.r.t. $\boldsymbol{\lambda}$ at $\bar{\boldsymbol{\lambda}}$.*

Proof Since (5.1) is the Lagrangian dual problem of (3.9), we apply Lemma 1 and the result follows. ∎

From Assumption 8, for each $\mathbf{x} \in \mathcal{X}$ and $k \in \{1, \ldots, K\}$,

$$G_k(\mathbf{x}) := \sup_{\boldsymbol{\lambda} \in \mathbb{R}^{l_2}} \{\mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x})\}. \tag{5.4}$$

We make the following additional assumption on the knowledge of the bounds on the optimal Lagrange multipliers and the optimum value of the Lagrangian. Note that the existence of these bounds is implied by Assumption 8.

Assumption 12 *There exists known constants $R_{\boldsymbol{\lambda}} > 1$, $\bar{s}$ and $\underline{s}$ such that, for $\forall \mathbf{x} \in \mathcal{X}$ and $k \in \{1, \ldots, K\}$, the optimal objective value of (5.4) is in $[\underline{s}, \bar{s}]$ and the interestion of the optimal solution of (5.4) and $B_{\boldsymbol{\lambda}}(0, R_{\boldsymbol{\lambda}})$ is nonempty.*

The following lemma guarantees the polynomial time solvability of (5.4).

Lemma 2 *Suppose that Assumption 12 is satisfied. Given $\mathbf{x} \in \mathcal{X}$ and $k \in \{1, \ldots, K\}$, we can find $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^{l_2}$, so that $\mathcal{L}_k(\bar{\boldsymbol{\lambda}}, \mathbf{x}) \geq G_k(\mathbf{x}) - \epsilon$ in polynomial time.*

Proof Given $\delta > 0$, denote $\hat{\delta} = \frac{\delta}{2}$ and consider the problem

$$\Theta_k(\mathbf{x}) := \min \ -s \tag{5.5a}$$

$$\text{s.t. } s \leq \mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x}) \tag{5.5b}$$

$$\underline{s} \leq s \leq \bar{s} \tag{5.5c}$$

$$\boldsymbol{\lambda} \in B_{\boldsymbol{\lambda}}(0, R_{\boldsymbol{\lambda}}). \tag{5.5d}$$

Since $\mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x})$ is a concave function w.r.t. $\boldsymbol{\lambda}$, the feasible region of (5.5) is convex. From Assumption 12, (5.5) is equivalent to (5.4). Let $C_k(\mathbf{x}) := \{(s; \boldsymbol{\lambda}) : s \leq \mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x}), \underline{s} \leq s \leq \bar{s}, \lambda \in B_{\boldsymbol{\lambda}}(0, R_{\boldsymbol{\lambda}})\}$ be the feasible region of (5.5). For $(\hat{s}; \hat{\boldsymbol{\lambda}}) \notin B_{s,\boldsymbol{\lambda}}(C_k(\mathbf{x}), \delta)$, either:

(i) $\hat{s} - \delta > \mathcal{L}_k(\hat{\boldsymbol{\lambda}}, \mathbf{x})$ or

(ii) $(\hat{s}, \hat{\boldsymbol{\lambda}})$ violates at least one of (5.5c) and (5.5d).

If $(\hat{s}, \hat{\boldsymbol{\lambda}})$ satisfies none of (i) and (ii), i.e., $\hat{s} - \delta \leq \mathcal{L}_k(\hat{\boldsymbol{\lambda}}, \mathbf{x})$ and (5.5c) and (5.5d) are satisfied, then obviously, $(\hat{s}, \hat{\boldsymbol{\lambda}}) \in B_{s,\boldsymbol{\lambda}}(C_k(\mathbf{x}), \delta)$. Let $\mathcal{L}_k^{\hat{\delta}}(\hat{\boldsymbol{\lambda}}, \mathbf{x})$ be the $\hat{\delta}$-precision value of $\mathcal{L}_k(\hat{\boldsymbol{\lambda}}, \mathbf{x})$ evaluated from Proposition 2 in polynomial time. $\mathcal{L}_k^{\bar{\delta}}(\hat{\boldsymbol{\lambda}}, \mathbf{x}) < \hat{s}$ because $\mathcal{L}_k(\hat{\boldsymbol{\lambda}}, \mathbf{x}) - \hat{\delta} \leq \mathcal{L}_k^{\hat{\delta}}(\hat{\boldsymbol{\lambda}}, \mathbf{x}) \leq \mathcal{L}_k(\hat{\boldsymbol{\lambda}}, \mathbf{x}) + \hat{\delta}$. For condition (ii), it is trivial to check the exact feasibility of $(\hat{s}, \hat{\boldsymbol{\lambda}})$ in polynomail time. Therefore, we get an oracle as:

(1) $\mathcal{L}_k^{\hat{\delta}}(\hat{\boldsymbol{\lambda}}, \mathbf{x}) < \hat{s}$;

(2) $\hat{s} < \underline{s}$ or $\hat{s} > \bar{s}$;

(3) $\hat{\boldsymbol{\lambda}} \notin B_{\boldsymbol{\lambda}}(0, R_{\boldsymbol{\lambda}})$.

Any $(\hat{s}, \hat{\boldsymbol{\lambda}}) \notin B_{s,\boldsymbol{\lambda}}(C_k(\mathbf{x}), \delta)$ will satisfy at least one of (1)-(3). Consequently, if $(\hat{s}, \hat{\boldsymbol{\lambda}})$ passes the oracle (1)-(3), $(\hat{s}, \hat{\boldsymbol{\lambda}}) \in B_{s,\boldsymbol{\lambda}}(C_k(\mathbf{x}), \delta)$. Now, we prove that for a given $(\hat{s}, \hat{\boldsymbol{\lambda}})$ not passing the oracle (1)-(3), we can generate a δ-cut, i.e, we can find a vector $(d_s; \mathbf{d}_{\boldsymbol{\lambda}})$ such that $(d_s; \mathbf{d}_{\boldsymbol{\lambda}})^T(\hat{s}, \hat{\boldsymbol{\lambda}}) < (d_s; \mathbf{d}_{\boldsymbol{\lambda}})^T(s, \boldsymbol{\lambda}) + \delta$ for $\forall (s, \boldsymbol{\lambda}) \in C_k(\mathbf{x})$. For a given $(\hat{s}, \hat{\boldsymbol{\lambda}})$, if (1) is satisfied, then according to Proposition 3, we can generate a $\hat{\delta}$-supergradient of $\mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x})$ at $\hat{\boldsymbol{\lambda}}$. Let $\mathbf{g}$ denote this $\hat{\delta}$-supergradient. Since

$$\mathcal{L}_k(\boldsymbol{\lambda}, \mathbf{x}) \leq \mathcal{L}_k(\hat{\boldsymbol{\lambda}}, \mathbf{x}) + \mathbf{g}^T(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}) + \hat{\delta} \text{ for } \forall \boldsymbol{\lambda} \in \mathbb{R}^{l_2},$$

inequality $s \leq \mathcal{L}_k(\hat{\boldsymbol{\lambda}}, \mathbf{x}) + \mathbf{g}^T(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}) + \hat{\delta}$ is valid for $\forall \boldsymbol{\lambda} \in \mathbb{R}^{l_2}$. We combine this inequality with $\mathcal{L}_k^{\hat{\delta}}(\hat{\boldsymbol{\lambda}}, \mathbf{x}) < \hat{s}$ and $\mathcal{L}_k(\hat{\boldsymbol{\lambda}}, \mathbf{x}) - \hat{\delta} \leq \mathcal{L}_k^{\hat{\delta}}(\hat{\boldsymbol{\lambda}}, \mathbf{x})$, and get a valid inequality:

$$s \leq \hat{s} + \mathbf{g}^T(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}) + 2\hat{\delta} \text{ for } \forall \boldsymbol{\lambda} \in \mathbb{R}^{l_2}.$$

It implies that:

$$(-1; \mathbf{g})^T(\hat{s}; \hat{\boldsymbol{\lambda}}) \leq (-1; \mathbf{g})^T(s; \boldsymbol{\lambda}) + \delta \text{ for } \forall (s; \boldsymbol{\lambda}) \in C_k(\mathbf{x}).$$

Obviously, $||(-1; \mathbf{g})||_\infty \geq 1$. If (2) is satisfied, either $\hat{s} < \underline{s}$ or $\hat{s} > \bar{s}$. The valid separating inequality for the first case is: $(1; \mathbf{0})^T(\hat{s}; \hat{\boldsymbol{\lambda}}) < (1; \mathbf{0})^T(s; \boldsymbol{\lambda})$ and for the second case is: $(-1; \mathbf{0})^T(\hat{s}; \hat{\boldsymbol{\lambda}}) < (-1; \mathbf{0}]^T(s; \boldsymbol{\lambda})$ for $\forall (s; \boldsymbol{\lambda}) \in C_k(\mathbf{x})$. If (3) is satisfied, a valid separating inequality is: $(0; -\gamma\hat{\boldsymbol{\lambda}})^T(\hat{s}; \hat{\boldsymbol{\lambda}}) < (0; -\gamma\hat{\boldsymbol{\lambda}})^T(\hat{s}; \boldsymbol{\lambda})$ for $\forall (s; \boldsymbol{\lambda}) \in C_k(\mathbf{x})$, where $\gamma > 0$ is a constant that ensures $\left|\left|(0, -\gamma\hat{\boldsymbol{\lambda}})\right|\right|_\infty \geq 1$. Let $\hat{\epsilon} = \frac{2}{3}\epsilon$ and s^* be an optimal solution of $\max_{(s,\boldsymbol{\lambda}) \in C_k(\mathbf{x})} s$. From Theorem 3 we have a $(\hat{s}; \hat{\boldsymbol{\lambda}}) \in B_{s,\boldsymbol{\lambda}}(C_k(\mathbf{x}), \epsilon)$ satisfying the oracle (1)-(3) with some $\delta \leq \hat{\epsilon}$ in polynomial time in $\log(\frac{1}{\epsilon})$, such that $\hat{s} \geq s^* - \hat{\epsilon}$. According to the definition of $C_k(\mathbf{x})$, we know that s^* is the optimal objective value of (5.4). On the other hand, since $(\hat{s}; \hat{\boldsymbol{\lambda}})$ satisfies the oracle (1)-(3) with some $\delta < \hat{\epsilon}$, we have that $s^* - \hat{\epsilon} \leq \hat{s} \leq \mathcal{L}_k^{\hat{\delta}}(\hat{\boldsymbol{\lambda}}, \mathbf{x}) \leq \mathcal{L}_k(\hat{\boldsymbol{\lambda}}, \mathbf{x}) + \hat{\delta}$. Therefore, $s^* < \mathcal{L}_k(\hat{\boldsymbol{\lambda}}, \mathbf{x}) + \frac{3}{2}\hat{\epsilon} = \mathcal{L}_k(\hat{\boldsymbol{\lambda}}, \mathbf{x}) + \epsilon$. We conclude that (5.4) can be solved to any precision ϵ in polynomial time. ∎

According to Proposition 2, for $\forall \mathbf{x} \in \mathcal{X}$ and $k \in \{1, \ldots, K\}$, the second stage problem $G_k(\mathbf{x})$ can be solved to any precision in polynomial time. The next lemma claims that the recourse function $G(\mathbf{x})$ can be evaluated to any precision in polynomial time.

Lemma 3 *Suppose that Assumption* 12 *is satisfied. Let* $\mathbf{x} \in \mathcal{X}$*, and the recourse function* $G(\mathbf{x})$ *be defined in* (1.2)*. Then, (i)* $G(\mathbf{x})$ *can be evaluated to any precision* ϵ*; (ii) an* ϵ*-subgradient of* $G(\mathbf{x})$ *can be obtained in time polynomial in* $\log(\frac{1}{\epsilon})$ *and* K*.*

Proof For a given $\epsilon > 0$, from Lemma 2, $G_k(\mathbf{x})$ can be obtained to $\hat{\epsilon} = \frac{\epsilon}{2}$-precision in polynomial time. Let $s_k\hat{\epsilon}, \boldsymbol{\lambda}_k^{\hat{\epsilon}}$ be an $\hat{\epsilon}$-optimal solution of (5.4). Since $-\hat{\epsilon} < G_k(\mathbf{x}) - s_k^{\hat{\epsilon}} < \hat{\epsilon}$, $s_k^{\hat{\epsilon}} + \hat{\epsilon} \geq \mathcal{L}_k(\boldsymbol{\lambda}_k^{\hat{\epsilon}}, \mathbf{x}) \geq s_k^{\hat{\epsilon}} - \hat{\epsilon}$, we see that $-\epsilon < G_k(\mathbf{x}) - \mathcal{L}_k(\boldsymbol{\lambda}_k^{\hat{\epsilon}}, \mathbf{x}) < \epsilon$. We can apply Theorem 5 to conclude that $-\mathbf{T}_k^T \boldsymbol{\lambda}_k^{\hat{\epsilon}}$ is an ϵ-subgradient of $G_k(\mathbf{x})$ at $\mathbf{x}$. Therefore, with $G_k(\mathbf{x}) := \sum_{k=1}^K \pi_k G_k(\mathbf{x})$, we have:

$$-\epsilon < G(\mathbf{x}) - \sum_{k=1}^{K} \pi_k s_k^{\hat{\epsilon}} < \epsilon,$$

which implies that $G(\mathbf{x})$ can be evaluated to ϵ-precision in polynomial time. From Theorem 1, $-\sum_{k=1}^K \pi_i \mathbf{T}_k^T \boldsymbol{\lambda}_k^{\hat{\epsilon}}$ is an ϵ-subgradient of $G(\cdot)$ at $\mathbf{x}$. ∎

The following Lemma shows that the feasibility of the semi-infinite constraint (5.6b) can be verified to any precision in polynomial time.

Lemma 4 (Delage and Ye 2010). *Assume the support set* $\mathscr{S} \subseteq \mathbb{R}^m$ *is convex and compact, and it is equipped with an oracle that for any* $\boldsymbol{\xi} \in \mathbb{R}^m$ *can either confirm that* $\boldsymbol{\xi} \in \mathscr{S}$ *or provide a hyperplane that separates* $\boldsymbol{\xi}$ *from* $\mathscr{S}$ *in time polynomial in* m*. Let function* $h(\mathbf{x}, \boldsymbol{\xi})$ *be concave in* $\boldsymbol{\xi}$ *in time polynomial in* m*. Then, for any* $\mathbf{x}$*,* $\mathbf{Y} \succeq 0$*, and* $\mathbf{y}$*, one can find a solution* $\boldsymbol{\xi}^*$ *that is* ϵ*-optimal with respect to the problem*

$$\max_{t, \boldsymbol{\xi}} \; t \tag{5.6a}$$

$$\text{s.t. } t \leq h(\mathbf{x}, \boldsymbol{\xi}) - \boldsymbol{\xi}^T \mathbf{Y} \boldsymbol{\xi} - \boldsymbol{\xi}^T \mathbf{y} \tag{5.6b}$$

$$\boldsymbol{\xi} \in \mathscr{S} \tag{5.6c}$$

in time polynomial in $\log(\frac{1}{\epsilon})$ *and the dimension of* $\boldsymbol{\xi}$*.* ∎

The next theorem shows the solvability of the two-stage problem (3.7)-(3.9) with recourse function $G(\mathbf{x})$.

Theorem 6 *Let Assumptions* 1-12 *be satisfied. Problem* (3.7)-(3.9) *can be solved to any precision* ϵ *in time polynomial in* $\log(\frac{1}{\epsilon})$ *and the size of the problem* (3.7)-(3.9).

Proof We want to apply Theorem 3 to show the polynomial solvability. The proof is divided into 5 steps. The first step is to verify that the feasible region is convex. Secondly, we prove the existence of an optimal solution of problem (3.7)-(3.9). In step 3 and 4, we establish the weak feasibility and weak separation oracles. We then verify the polynomial solvability of (3.7)-(3.9) by applying Theorem 3 in step 5.

Step 1. Verification of the Convexity of the Feasible Region

Let $\mathbf{z} := (\mathbf{x}, \mathbf{Y}_1, \mathbf{y}_1, y_1^0, t_1, \tau)$ and rewrite (3.7) as:

$$\min_{\mathbf{z}} \; y_1^0 + t_1 + \tau \tag{5.7a}$$
$$s.t. \; \tau \geq G(\mathbf{x}) \tag{5.7b}$$
$$y_1^0 \geq \rho_1(\mathbf{x}, \mathbf{p}) - \mathbf{p}^T \mathbf{Y}_1 \mathbf{p} - \mathbf{p}^T \mathbf{y}_1, \; \forall \mathbf{p} \in \mathscr{S}_1 \tag{5.7c}$$
$$t_1 \geq (\beta_1 \mathbf{\Sigma}_1 + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^T) \bullet \mathbf{Y}_1 + \boldsymbol{\mu}_1^T \mathbf{y}_1 + \sqrt{\alpha_1} \left|\left| \mathbf{\Sigma}_1^{\frac{1}{2}} (\mathbf{y}_1 + 2\mathbf{Y}_1 \boldsymbol{\mu}_1) \right|\right| \tag{5.7d}$$
$$\mathbf{Y}_1 \succeq 0 \tag{5.7e}$$
$$\mathbf{x} \in \mathcal{X}. \tag{5.7f}$$

From Proposition 2, for $\forall k \in \{1, \ldots, K\}$, $G_k(\mathbf{x})$ can be evaluated to any precision in polynomial time. Since $G_k(\mathbf{x})$ is a convex function w.r.t. $\mathbf{x}$, $G(\mathbf{x})$ is a convex function w.r.t. $\mathbf{x}$. It implies that (5.7b) is a convex contraint. According to Assumption 5, (5.7e) and (5.7c) are convex, (5.7d) is a second order cone constraint, which is convex. Constraints (5.7e) and (5.7f) are obviously convex. Therefore, the feasible region of (5.7) is convex.

Step 2. Existence of an Optimal Solution of (3.7)-(3.9).

Let $\bar{\mathbf{x}} \in \mathcal{X}$ and $\bar{\mathbf{z}} := (\bar{\mathbf{x}}, \bar{\mathbf{Y}}_1, \bar{\mathbf{y}}_1, \bar{y}_1^0, \bar{t}_1, \bar{\tau})$ be defined as: $\bar{\tau} = G(\bar{\mathbf{x}})$, $\bar{\mathbf{Y}} = \mathbf{I}$, $\bar{\mathbf{y}} = 0$, $\bar{y}_1^0 = \sup_{\mathbf{p} \in \mathscr{S}_1} \{\rho_1(\bar{\mathbf{x}}, \mathbf{p}) - \mathbf{p}^T \bar{\mathbf{Y}}_1 \mathbf{p}\}$, $\bar{t}_1 = trace(\beta_1 \mathbf{\Sigma}_1 + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^T) + 2\sqrt{\alpha_1} \left|\left| \mathbf{\Sigma}_1^{\frac{1}{2}} \boldsymbol{\mu}_1 \right|\right|$. Then $\bar{\mathbf{z}}$ is a feasible solution of (5.7). Note that $\bar{y}_1^0$ exists because $\mathscr{S}_1$ is compact. Therefore, the feasible region of (5.7) is nonempty. On the other hand, from Theorem 1 and Assumptions 1-5, the set of optimal solutions of the problem:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{Y}_1, \mathbf{y}_1, y_1^0, t_1} \quad & y_1^0 + t_1 \\ s.t. \quad & y_1^0 \geq \rho_1(\mathbf{x}, \mathbf{p}) - \mathbf{p}^T \mathbf{Y}_1 \mathbf{p} - \mathbf{p}^T \mathbf{y}_1, \; \forall \mathbf{p} \in \mathscr{S}_1, \\ & t_1 \geq (\beta_1 \mathbf{\Sigma}_1 + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^T) \bullet \mathbf{Y}_1 + \boldsymbol{\mu}_1^T \mathbf{y}_1 + \sqrt{\alpha_1} \left|\left| \mathbf{\Sigma}_1^{\frac{1}{2}} (\mathbf{y}_1 + 2\mathbf{Y}_1 \boldsymbol{\mu}_1) \right|\right|, \\ & \mathbf{Y}_1 \succeq 0, \; \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{5.8}$$

is nonempty. Let us assume that the optimal objective value of (5.8) equals γ_1. From Lemma 2, the recourse function $G(\mathbf{x})$ is finite for $\forall \mathbf{x} \in \mathcal{X}$. Since $\mathcal{X}$ is compact, $\gamma_2 := \min_{\mathbf{x} \in \mathcal{X}} G(\mathbf{x})$ is finite. Since the optimal objective value of (5.7) is no less than $\gamma_1 + \gamma_2$, we conclude that (5.7) has a finite objective value and the set of optimal solutions is nonempty.

Step 3. Establishment of the Weak Feasibility Oracle

According to Assumption 3, we know that $B_{\mathbf{x}}(\mathbf{x}_0, r_0^1) \subset \mathcal{X} \subset B_{\mathbf{x}}(0, R_0^1)$. Let $\underline{\mathbf{x}} = \mathbf{x}_0$. $\underline{\mathbf{Y}}_1 = \mathbf{I}$ and $\underline{\mathbf{y}}_1 = 0$. Let $0 < r_0 < r_0^1$ be a constant such that $\underline{\mathbf{Y}}_1 + \mathbf{S} \succeq 0$ for $\forall \, ||\mathbf{S}||_F \leq r_0$, where

$||\cdot||_F$ is the Frobenius norm of matrices. Let

$$\underline{y}_1^0 = \sup_{||\mathbf{Y}_1-\mathbf{I}||_F\le r_0, ||\mathbf{y}_1||\le r_0, ||\mathbf{x}-\mathbf{x}_0||\le r_0, \mathbf{p}\in\mathscr{S}_1} \{\rho_1(\mathbf{x},\mathbf{p}) - \mathbf{p}^T\mathbf{Y}_1\mathbf{p} - \mathbf{p}^T\mathbf{y}_1\} + r_0$$

$$\underline{t}_1 = \sup_{||\mathbf{Y}_1-\mathbf{I}||_F\le r_0, ||\mathbf{y}_1||<r_0} \{(\beta_1\mathbf{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T)\bullet\mathbf{Y}_1 + \boldsymbol{\mu}_1^T\mathbf{y}_1 + \sqrt{\alpha_1}\left|\left|\mathbf{\Sigma}_1^{\frac{1}{2}}(\mathbf{y}_1 + 2\mathbf{Y}_1\boldsymbol{\mu}_1)\right|\right|\} + r_0$$

$$\underline{\tau} = \sup_{\mathbf{x}\in B_{\mathbf{x}}(\mathbf{x}_0,r_0)} G(\mathbf{x}) + r_0.$$

$$\underline{\mathbf{z}} := (\underline{\mathbf{x}}, \underline{\mathbf{Y}}_1, \underline{\mathbf{y}}_1, \underline{\tau}, \underline{y}_1^0, \underline{t}_1).$$

Then, the feasible region of problem (5.7) contains the ball $B_{\mathbf{z}}(\underline{\mathbf{z}}, r_0)$. On the other hand, let $R_0 > r_0 + ||\underline{\mathbf{z}}||$ be a constant such that the intersection of the set of optimal solutions of problem (5.7) and the ball $B_{\mathbf{z}}(0, R_0)$ is nonempty. Consequently, solving (5.7) is equivalent to (5.7) with the additional constraint $\mathbf{z}\in B_{\mathbf{z}}(0,R_0)$. Let the feasible region of problem (5.7) with this additional constraint be C. From the above discussion, we have $B_{\mathbf{z}}(\underline{\mathbf{z}}, r_0) \subset C \subset B_{\mathbf{z}}(0, R_0)$. Given $\delta > 0$, let $\hat{\delta} = \frac{\delta}{2}$. Let $\hat{\mathbf{z}} := (\hat{\mathbf{x}}, \hat{\mathbf{Y}}_1, \hat{\mathbf{y}}_1, \hat{\tau}, \hat{y}_1^0, \hat{t}_1) \notin B_{\mathbf{z}}(C, \delta)$. The point $\hat{\mathbf{z}}$ satisfies at least one of the following three conditions.

(i) $\hat{\tau} + \hat{\delta} < G(\hat{\mathbf{x}})$;

(ii) $\hat{y}_1^0 + \hat{\delta} < \sup_{\mathbf{p}\in\mathscr{S}_1}\{\rho_1(\hat{\mathbf{x}},\mathbf{p}) - \mathbf{p}^T\hat{\mathbf{Y}}_1\mathbf{p} - \mathbf{p}^T\hat{\mathbf{y}}_1\}$;

(iii) $\hat{\mathbf{z}}$ at least violates one of (5.7d)-(5.7f).

If none of the conditions (i)-(iii) is satisfied, then $\hat{\mathbf{z}} \in B_{\mathbf{z}}(C,\delta)$. If (i) is satisfied, since $G(\mathbf{x})$ can be evaluated to precision $\frac{\epsilon}{2}$ in polynomial time for $\forall \mathbf{x}\in\mathcal{X}$, we have that $\hat{\tau} < G^{\hat{\delta}}(\hat{\mathbf{x}})$, where $G^{\hat{\delta}}(\hat{\mathbf{x}})$ is the $\hat{\delta}$-optimal objective value of $G(\hat{\mathbf{x}})$ according to Proposition 2. Assume that (ii) is satisfied. According to Lemma 4, $\sup_{\mathbf{p}\in\mathscr{S}_1}\{\rho_1(\hat{\mathbf{x}},\mathbf{p}) - \mathbf{p}^T\hat{\mathbf{Y}}_1\mathbf{p} - \mathbf{p}^T\hat{\mathbf{y}}_1\}$ can be found to $\hat{\delta}$ precision polynomially. Let $\hat{\mathbf{p}}$ be the corresponding $\hat{\delta}$ optimal solution. Then condition (ii) implies that $\hat{y}_1^0 < \rho_1(\hat{\mathbf{x}},\hat{\mathbf{p}}) - \hat{\mathbf{p}}^T\hat{\mathbf{Y}}_1\hat{\mathbf{p}} - \hat{\mathbf{p}}^T\hat{\mathbf{y}}_1$. We have an oracle system as:

$$\hat{\tau} < G^{\hat{\delta}}(\hat{\mathbf{x}}); \tag{5.9}$$

$$\hat{y}_1^0 < \rho_1(\hat{\mathbf{x}},\hat{\mathbf{p}}) - \hat{\mathbf{p}}^T\hat{\mathbf{Y}}_1\hat{\mathbf{p}} - \hat{\mathbf{p}}^T\hat{\mathbf{y}}_1; \tag{5.10}$$

$$\hat{t}_1 < (\beta^1\mathbf{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T)\bullet\hat{\mathbf{Y}}_1 + \boldsymbol{\mu}_1^T\hat{\mathbf{y}}_1 + \sqrt{\alpha_1}\left|\left|\mathbf{\Sigma}_1^{\frac{1}{2}}(\hat{\mathbf{y}}_1 + 2\hat{\mathbf{Y}}_1\boldsymbol{\mu}_1)\right|\right|; \tag{5.11}$$

$$\hat{\mathbf{Y}}_1 \not\succeq 0; \tag{5.12}$$

$$\hat{\mathbf{x}} \notin \mathcal{X}. \tag{5.13}$$

We need to show that the system (5.9)-(5.13) can be verified in polynomial time. Condition (5.9) can be verified in polynomial time by using Lemma 3. Condition (5.10) can be verified in polynomial time by using Lemma 4. Condition (5.11) can be verified by verifing the feasibility of (5.7d). Condition (5.12) can be verified using matrix factorization in $O(m_1^3)$ arithmetic operations. Condition (5.13) can be verified in polynomial time according to Assumption 3. In addition from the above discussion, if $\hat{\mathbf{z}}$ does not satisfy any of (5.9)-(5.13), then $\hat{\mathbf{z}} \in B_{\mathbf{z}}(C,\delta)$.

Step 4. Establishment of the Weak Separation Oracle

In Step 3, we showed that the oracle system (5.9)-(5.13) can find any $\hat{\mathbf{z}} \notin B_{\mathbf{z}}(C, \delta)$ and can be verified in polynomial time. If any of (5.9)-(5.13) are satisfied, we will prove that we can generate an inequality which separates $\hat{\mathbf{z}}$ from the feasibility region C with δ-tolerence, i.e. satisfy the condition 2 described in Theorem 3. Assume that a point $\hat{\mathbf{z}}$ is given. If (5.9) is satisfied, from Lemma 3 we can obtain a $\hat{\delta}$-subgradient of $G(\mathbf{x})$ at $\hat{\mathbf{x}}$. Let us denote this $\hat{\delta}$-subgradient as $\mathbf{g}$. Since $G(\mathbf{x}) \geq G(\hat{\mathbf{x}}) + \mathbf{g}^T(\mathbf{x} - \hat{\mathbf{x}}) - \hat{\delta}$ for $\forall \mathbf{x} \in \mathcal{X}$, we generate a valid inequality: $\tau \geq G(\hat{\mathbf{x}}) + \mathbf{g}^T(\mathbf{x} - \hat{\mathbf{x}}) - \hat{\delta}$. Combining with $G^{\hat{\delta}}(\hat{\mathbf{x}}) - \hat{\delta} \leq G(\hat{\mathbf{x}}) \leq G^{\hat{\delta}}(\hat{\mathbf{x}}) + \hat{\delta}$ and $G^{\hat{\delta}}(\hat{\mathbf{x}}) > \hat{\tau}$, we get the separating hyperplane (5.14):

$$(-\mathbf{g}; 0; 0; 1; 0; 0)^T vec(\mathbf{z}) + \delta \geq (-\mathbf{g}; 0; 0; 1; 0; 0]^T vec(\hat{\mathbf{z}}) \tag{5.14}$$

for $\forall \mathbf{z} \in C$, where $vec(\mathbf{z}) := (\mathbf{x}; vec(\mathbf{Y}_1); \mathbf{y}_1; y_1^0; t_1; \tau)$ and $vec(\cdot)$ vectorizes a matrix. It is clear that: $||(\mathbf{g}; 0; 0; 1; 0; 0)||_\infty \geq 1$.

If (5.10) is satisfied, assume $\hat{\mathbf{p}}$ be the $\hat{\delta}$-optimal solution of problem (5.6) w.r.t. $\hat{\mathbf{z}}$. Then, we can generate the separating hyperplane:

$$-(\nabla_{\mathbf{x}}\rho_1(\hat{\mathbf{x}}, \hat{\mathbf{p}}); vec(\hat{\mathbf{p}}\hat{\mathbf{p}}^T); \hat{\mathbf{p}}; 0; 1; 0)^T vec(\mathbf{z}) \geq -(\nabla_{\mathbf{x}}\rho_1(\hat{\mathbf{x}}, \hat{\mathbf{p}}); vec(\hat{\mathbf{p}}\hat{\mathbf{p}}^T); \hat{\mathbf{p}}; 0; 1; 0)^T vec(\hat{\mathbf{z}}), \tag{5.15}$$

for $\forall \mathbf{z} \in C$, where $\nabla_{\mathbf{x}}\rho_1(\mathbf{x}, \mathbf{p})$ is a subgradient of $\rho_1(\mathbf{x}, \mathbf{p})$ w.r.t. $\mathbf{x}$ (Assumption 4). Note that: $\left|\left|(\nabla_{\mathbf{x}}\rho_1(\hat{\mathbf{x}}, \hat{\mathbf{p}}); vec(\hat{\mathbf{p}}\hat{\mathbf{p}}^{\hat{\delta}T}); \hat{\mathbf{p}}; 0; 1; 0)\right|\right|_\infty \geq 1$.

If (5.11) is satisfied, a valid inequality can be generated as: $(\beta_1\mathbf{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T + \hat{\mathbf{G}}) \bullet \mathbf{Y}_1 + (\boldsymbol{\mu}_1 + \hat{\mathbf{g}})^T\mathbf{y}_1 - t_1 \leq \hat{\mathbf{g}}^T\hat{y}_1 + \hat{\mathbf{G}} \bullet \hat{\mathbf{Y}}_1 - g(\hat{\mathbf{Y}}_1, \hat{\mathbf{y}}_1)$, where $\hat{\mathbf{g}} = \nabla_{\mathbf{y}_1} g(\hat{\mathbf{Y}}_1, \hat{\mathbf{y}}_1)$, $\hat{G} = mat(\nabla_{\mathbf{Y}_1} g(\hat{\mathbf{Y}}_1, \hat{\mathbf{y}}_1))$, $g(\mathbf{Y}_1, \mathbf{y}_1) := \left|\left|\mathbf{\Sigma}_1^{\frac{1}{2}}(\mathbf{y}_1 + 2\mathbf{Y}_1\boldsymbol{\mu}_1)\right|\right|$, and $\nabla_{\mathbf{Y}_1} g(\mathbf{Y}_1, \mathbf{y}_1)$; $\nabla_{\mathbf{y}_1} g(\mathbf{Y}_1, \mathbf{y}_1)$ are the gradients of $g(\mathbf{Y}_1, \mathbf{y}_1)$ in $\mathbf{Y}_1$ and $\mathbf{y}_1$ respectively. Note that $mat(\cdot)$ puts a vector to a symmetric square matrix. It implies that:

$$\begin{aligned}&-(0; vec(\beta_1\mathbf{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T + \hat{\mathbf{G}}); (\boldsymbol{\mu}_1 + \hat{\mathbf{g}}); 0; 1)^T vec(\hat{\mathbf{z}}) < \\ &-(0; vec(\beta_1\mathbf{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T + \hat{\mathbf{G}}); (\boldsymbol{\mu}_1 + \hat{\mathbf{g}}); 0; 1)^T vec(\mathbf{z})\end{aligned}$$

for $\forall \mathbf{z} \in C$. Obviously,

$$\left|\left|(0; vec(\beta_1\mathbf{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T + \hat{\mathbf{G}}); (\boldsymbol{\mu}_1 + \hat{\mathbf{g}}); 0; 1)\right|\right|_\infty \geq 1.$$

If (5.12) is satisfied, a separating hyperplane can be generated based on the eigenvector corresponding to the lowest eigenvalue. The vector to represent this nonzero separating hyperplane can be scaled to satisfy the requirement $||\cdot||_\infty \geq 1$. If (5.13) is satisfied, a separating hyperplane with $||\cdot||_\infty \geq 1$ can be generated in polynoial time according to Assumption 3.

Step 5. Verification of the Polynomial Solvability

According to the analysis in steps 1-4, we can apply Lemma 3 to conclude that for a given $\epsilon > 0$ and $\hat{\epsilon} = \frac{\epsilon}{3}$, we can find a $\hat{\mathbf{z}} := (\hat{\mathbf{x}}, \hat{\mathbf{Y}}_1, \hat{\mathbf{y}}_1, \hat{y}_1^0, \hat{t}_1, \hat{\tau})$ in time polynomial in $\log(\frac{1}{\epsilon})$, such that $\hat{\mathbf{z}}$ satisfies the oracle (5.9)-(5.13) with some $\delta < \hat{\epsilon}$ and $\hat{y}_1^0 + \hat{t}_1 + \hat{\tau} < \eta^* + \hat{\epsilon}$, where η^* is the true optimal value of the problem (3.7). Since $\hat{\mathbf{z}}$ satisfies the oracle (5.9)-(5.13) with $\delta < \hat{\epsilon}$, the solution $\hat{\mathbf{z}}^* = (\hat{\mathbf{x}}, \hat{\mathbf{Y}}_1, \hat{\mathbf{y}}_1, \hat{y}_1^0 + \hat{\epsilon}, \hat{t}_1, \hat{\tau} + \hat{\epsilon}) \in C$ and the objective value associated with $\hat{\mathbf{z}}^*$ is no greater than $\eta^* + 3\hat{\epsilon} = \eta^* + \epsilon$.

In summary, we have shown that problem (3.7)-(3.9) can be solved to any precision in polynomial time. ∎

Remark: Delage and Ye [5] use Lemma 4 to verify the feasibility of the constraint (5.6b). For infeasibility point, a separating hyperplane can be generated by using the optimal solution of (5.6). The assumption they make is that an exact solution of (5.6) can be found. Lemma 4 can only claim that the problem can be solved to ϵ-precision for the two-stage moment robust model. In the proof of Theorem 6, we show that it is enough to use the $\hat{\delta}$ optimal solution to verify the feasibility of (5.6) and generate a separating hyperplane.

6 A Two-Stage Moment Robust Portfolio Optimization Model

In this section we analyze a two-stage moment robust model with piecewise linear objectives. In Section 6.1 we show that when (1) the probability ambiguity sets $\mathscr{P}_1$ an $\mathscr{P}_{2,k}$ are described by (1.6)-(1.7) or exact moment information; (2) the support $\mathscr{S}_1$ and $\mathscr{S}_2$ are the full space $\mathbb{R}^{n_1}, \mathbb{R}^{n_2}$ or described by ellipsoids, the two stage moment robust problem can be reformulated as a semidefinite program. Note that the single stage moment robust model with piecewise linear objective and probability ambiguity set described by exact moments are discussed by Bertsimas et al. [2]. In Section 6.2 we study a two-stage moment robust portfolio optimization application with practical data. Our numerical results suggest that the two-stage modeling is effective when we have forecasting power.

6.1 A Two-Stage Moment Robust Optimization Model with Piecewise Linear Objectives

Consider the two-stage moment robust optimization model:

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}\in\mathscr{P}_1} \mathbb{E}_{\mathbb{P}}[U(\tilde{\mathbf{p}}^T\mathbf{x})] + G(\mathbf{x}), \tag{6.1}$$

$$G(\mathbf{x}) := \sum_{k=1}^{K} \pi_k G_k(\mathbf{x}), \tag{6.2}$$

$$G_k(\mathbf{x}) := \min_{\mathbf{w}_k\in\mathcal{W}_k(\mathbf{x})} \max_{\mathbb{P}\in\mathscr{P}_{2,k}} [U(\tilde{\mathbf{q}}^T\mathbf{w}_k)], \tag{6.3}$$

where $\mathcal{X}$ and $\mathcal{W}_k(\mathbf{x})$ are described by linear, second-order cone and semidefinite constraints, and the ambiguity sets $\mathscr{P}_1$ and $\mathscr{P}_{2,k}$ are defined in (1.6)-(1.7). The utility function $U(\cdot)$ is piecewise linear convex and defined as: $U(z) = \max_{i=1,\ldots,M}\{c_i z + d_i\}$, where c_k, d_k, $k = 1, \ldots, M$ are given. Let Assumption 1 and Assumption 6 be satisfied. By applying Theorem

1, problem (6.1)-(6.3) is equivalent to:

$$\begin{aligned}
&\min_{\mathbf{x},\mathbf{Y}_1,\mathbf{y}^1,y_1^0,t_1} \quad y_1^0 + t_1 + G(\mathbf{x}), &(6.4)\\
&\text{s.t. } y_1^0 \geq c_i\mathbf{p}^T\mathbf{x} + d_i - \mathbf{p}^T\mathbf{Y}_1\mathbf{p} - \mathbf{p}^T\mathbf{y}_1,\ \forall \mathbf{p} \in \mathscr{S}_1, i = 1,\dots,M,\\
&\qquad t_1 \geq (\beta_1\mathbf{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T) \bullet \mathbf{Y}_1 + \boldsymbol{\mu}_1^T\mathbf{y}_1 + \sqrt{\alpha_1}\left|\left|\mathbf{\Sigma}_1^{\frac{1}{2}}(\mathbf{y}_1 + 2\mathbf{Y}_1\boldsymbol{\mu}_1)\right|\right|,\\
&\qquad \mathbf{Y}_1 \succeq 0,\ \mathbf{x} \in \mathcal{X},
\end{aligned}$$

where

$$\begin{aligned}
&G(\mathbf{x}) := \sum_{k=1}^{K} \pi_k G_k(\mathbf{x}), &(6.5)\\
&G_k(\mathbf{x}) := \min_{\mathbf{w}_k,\mathbf{Y}_{2,k},\mathbf{y}_{2,k},y_{2,k}^0,t_{2,k}} \quad y_{2,k}^0 + t_{2,k}, &(6.6)\\
&\text{s.t. } y_{2,k}^0 \geq c_i\mathbf{q}^T\mathbf{w}_k + d_i - \mathbf{q}^T\mathbf{Y}_{2,k}\mathbf{q} - \mathbf{q}^T\mathbf{y}_{2,k},\ \forall \mathbf{q} \in \mathscr{S}^2, i = 1,\dots,M,\\
&\qquad t_{2,k} \geq (\beta_2\mathbf{\Sigma}_{2,k} + \boldsymbol{\mu}_{2,k}\boldsymbol{\mu}_{2,k}^T) \bullet \mathbf{Y}_{2,k} + \boldsymbol{\mu}_{2,k}^T\mathbf{y}_{2,k},\\
&\qquad\qquad + \sqrt{\alpha_2}\left|\left|\mathbf{\Sigma}_{2,k}^{\frac{1}{2}}(\mathbf{y}_{2,k} + 2\mathbf{Y}_{2,k}\boldsymbol{\mu}_{2,k})\right|\right|,\\
&\qquad \mathbf{Y}_{2,k} \succeq 0,\ \mathbf{w}_k \in \mathcal{W}_k(\mathbf{x}).
\end{aligned}$$

The piecewise utility function U can be understood as a convex utility on the linear objective $\mathbf{p}^T\mathbf{x}$ and $\mathbf{q}_k^T\mathbf{w}_k$. Now we discuss three subcases: (i) Assume that $\mathscr{S}_1$ and $\mathscr{S}_2$ are convex and compact; (ii)$\mathscr{S}_1 = \mathbb{R}^{n_1}$ and $\mathscr{S}_2 = \mathbb{R}^{n_2}$, where $n_1 = \dim(\mathbf{x})$ and $n_2 = \dim(\mathbf{w}_k)$; (iii) $\mathscr{S}_1$ and $\mathscr{S}_2$ are described by some ellipsoids, i.e. $\mathscr{S}_1 = \{\mathbf{p} : (\mathbf{p} - \mathbf{p}_0)^T\mathbf{Q}_1(\mathbf{p} - \mathbf{p}_0) \leq 1\}$, $\mathscr{S}_2 = \{\mathbf{q} : (\mathbf{q} - \mathbf{q}_0)^T\mathbf{Q}_2(\mathbf{q} - \mathbf{q}_0) \leq 1\}$, where $\mathbf{p}_0$, $\mathbf{Q}_1$, $\mathbf{q}_0$ and $\mathbf{Q}_2$ are given and matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ have at least one strictly positive eigenvalue. For case (i), problem (6.1)-(6.3) will be a special case of the general model analyzed in Section 5. For cases (ii) and (iii), we give two-stage semidefinite reformulations of (6.4)-(6.6). The next lemma from Delage and Ye [5] provide an equivalent reformulation of the semi-infinite constraints in (6.4)-(6.6).

Lemma 5 (Delage and Ye 2012 [5]). *The semi-infinite constraints*

$$\boldsymbol{\xi}^T\mathbf{Y}\boldsymbol{\xi} + \boldsymbol{\xi}^T\mathbf{y} + y_0 \geq c_i\boldsymbol{\xi}^T\mathbf{x} + d_i, \forall \boldsymbol{\xi} \in \mathscr{S}, i = 1,\dots,M, \tag{6.7}$$

can be reformulated as the following semidefinite constraints.

(1) *If $\mathscr{S} = \mathbb{R}^n$, we can reformulate* (6.7) *as:*

$$\begin{pmatrix} \mathbf{Y} & \frac{1}{2}(\mathbf{y} - c_i\mathbf{x}) \\ \frac{1}{2}(\mathbf{y} - c_i\mathbf{x})^T & y_0 - d_i \end{pmatrix} \succeq 0, \forall i = 1,\dots,M. \tag{6.8}$$

(2) *If $\mathscr{S} = \{\boldsymbol{\xi} : (\boldsymbol{\xi} - \boldsymbol{\xi}_0)^T\mathbf{\Theta}(\boldsymbol{\xi} - \boldsymbol{\xi}_0) \leq 1\}$, we can reformulate* (6.7) *as:*

$$\begin{pmatrix} \mathbf{Y} & \frac{1}{2}(\mathbf{y} - c_i\mathbf{x}) \\ \frac{1}{2}(\mathbf{y} - c_i\mathbf{x})^T & y_0 - d_i \end{pmatrix} \succeq \tau_i \begin{pmatrix} \mathbf{\Theta} & -\mathbf{\Theta}\boldsymbol{\theta}_0 \\ -\boldsymbol{\theta}_0^T\mathbf{\Theta} & \boldsymbol{\theta}_0^T\mathbf{\Theta}\boldsymbol{\theta}_0 - 1 \end{pmatrix}, \tau_i \geq 0, \forall i = 1,\dots,M, \tag{6.9}$$

where $\boldsymbol{\theta}_0$ is given and $\mathbf{\Theta} \succeq 0$ has at least one strictly positive eigenvalue.

By the following theorem applying Lemma 5 to (6.4)-(6.6), we can reformulate (6.1)-(6.3) as a two-stage semidefinite program.

Theorem 7 *Let Assumption* 1 *and Assumption* 6 *be satisfied. If* (*i*) $\mathscr{S}_1 = \mathbb{R}^{n_1}$ *and* $\mathscr{S}_2 = \mathbb{R}^{n_2}$; *or* (*ii*) $\mathscr{S}_1 = \{\mathbf{p} : (\mathbf{p}-\mathbf{p}_0)^T\mathbf{Q}_1(\mathbf{p}-\mathbf{p}_0) \le 1\}$, $\mathscr{S}_2 = \{\mathbf{q} : (\mathbf{q}-\mathbf{q}_0)^T\mathbf{Q}_2(\mathbf{q}-\mathbf{q}_0) \le 1\}$, *where* $\mathbf{p}_0$, $\mathbf{Q}_1$, $\mathbf{q}_0$ *and* $\mathbf{Q}_2$ *are given and matrices* $\mathbf{Q}_1$ *and* $\mathbf{Q}_2$ *have at least one strictly positive eigenvalue, then the two-stage moment robust problem* (6.1)-(6.3) *is equivalent to the two-stage semidefinite programming problem:*

$$
\begin{aligned}
&\min_{\mathbf{x},\mathbf{Y}_1,\mathbf{y}_1,y_1^0,t_1} \quad y_1^0 + t_1 + G(\mathbf{x}), && (6.10)\\
&\text{s.t.} \quad \begin{pmatrix} \mathbf{Y}_1 & \frac{\mathbf{y}_1 - c_i\mathbf{x}}{2} \\ \frac{(\mathbf{y}_1 - c_i\mathbf{x})^T}{2} & y_1^0 - d_i \end{pmatrix} \succeq \tau_i \mathbf{B}_1, \ \text{for } i = 1, \dots, M,\\
&\qquad t_1 \ge (\beta_1 \boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T) \bullet \mathbf{Y}_1 + \boldsymbol{\mu}_1^T\mathbf{y}_1 + \sqrt{\alpha_1}\left\|\boldsymbol{\Sigma}_1^{\frac{1}{2}}(\mathbf{y}_1 + 2\mathbf{Y}_1\boldsymbol{\mu}_1)\right\|,\\
&\qquad \mathbf{Y}_1 \succeq 0, \mathbf{x} \in \mathcal{X},
\end{aligned}
$$

where

$$
\begin{aligned}
&G(\mathbf{x}) := \sum_{k=1}^{K} \pi_k G_k(\mathbf{x}), && (6.11)\\
&G_k(\mathbf{x}) := \min_{\mathbf{w}_k,\mathbf{Y}_{2,k},\mathbf{y}_{2,k},y_{2,k}^0,t_{2,k}} \quad y_{2,k}^0 + t_{2,k}, && (6.12)\\
&\text{s.t.} \quad \begin{pmatrix} \mathbf{Y}_{2,k} & \frac{\mathbf{y}_{2,k} - c_i\mathbf{w}_k}{2} \\ \frac{(\mathbf{y}_{2,k} - \mathbf{w}_k)^T}{2} & y_{2,k}^0 - d_i \end{pmatrix} \succeq \tau_i \mathbf{B}_2 \ \text{for } i = 1, \dots, M,\\
&\qquad t_{2,k} \ge (\beta_2 \boldsymbol{\Sigma}_{2,k} + \boldsymbol{\mu}_{2,k}\boldsymbol{\mu}_{2,k}^T) \bullet \mathbf{Y}_{2,k} + \boldsymbol{\mu}_{2,k}^T\mathbf{y}_{2,k}\\
&\qquad\qquad + \sqrt{\alpha_2}\left\|\boldsymbol{\Sigma}_{2,k}^{\frac{1}{2}}(\mathbf{y}_{2,k} + 2\mathbf{Y}_{2,k}\boldsymbol{\mu}_{2,k})\right\|,\\
&\qquad \mathbf{Y}_{2,k} \succeq 0, \mathbf{w}_k \in \mathcal{X},
\end{aligned}
$$

where

(1) $\mathbf{B}_1 = \mathbf{B}_2 = 0$ if $\mathscr{S}_1 = \mathbb{R}^{n_1}$, $\mathscr{S}_2 = \mathbb{R}^{n_2}$;

(2) $\mathbf{B}_1 = \begin{pmatrix} \mathbf{Q}_1 & -\mathbf{Q}_1\mathbf{p}_0 \\ -\mathbf{p}_0^T\mathbf{Q}_1 & \mathbf{p}_0^T\mathbf{Q}_1\mathbf{p}_0 - 1 \end{pmatrix}$, $\mathbf{B}_2 = \begin{pmatrix} \mathbf{Q}_2 & -\mathbf{Q}_2\mathbf{q}_0 \\ -\mathbf{q}_0^T\mathbf{Q}_2 & \mathbf{q}_0^T\mathbf{Q}_2\mathbf{q}_0 - 1 \end{pmatrix}$, if $\mathscr{S}_1 = \{\mathbf{p} : (\mathbf{p} - \mathbf{p}_0)^T\mathbf{Q}_1(\mathbf{p} - \mathbf{p}_0) \le 1\}$, $\mathscr{S}_2 = \{\mathbf{q} : (\mathbf{q} - \mathbf{q}_0)^T\mathbf{Q}_2(\mathbf{q} - \mathbf{q}_0) \le 1\}$.

Equations (6.10)-(6.12) are standard two-stage linear semidefinite programming problem. Two-stage linear semidefinite programs can be solved using interior decomposition methods in [11, 12, 10]. The performance will be demonstrated in our context in Section 6.2.

Another interesting case of model (6.1)-(6.6) is to assume that the ambiguity sets $\mathscr{P}_1$ and $\mathscr{P}_{2,k}$ are described by the exact mean vector and covariance matrix, i.e. $\mathscr{P}_1$ and $\mathscr{P}_{2,k}$ are given as:

$$
\begin{aligned}
\mathscr{P}_1 :=& \{\mathbb{P} : \mathbb{P} \in \mathcal{M}_1, \mathbb{E}_{\mathbb{P}}[\mathbf{1}] = \mathbf{1}, \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{p}}] = \boldsymbol{\mu}_1, \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{p}}\tilde{\mathbf{p}}^T] = \boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T, \tilde{p} \in \mathscr{S}_1\}, && (6.13)\\
\mathscr{P}_{2,k} :=& \{\mathbb{P} : \mathbb{P} \in \mathcal{M}_2, \mathbb{E}_{\mathbb{P}}[\mathbf{1}] = \mathbf{1}, \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{q}}] = \boldsymbol{\mu}_{2,k}, \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{q}}\tilde{\mathbf{q}}^T] = \boldsymbol{\Sigma}_{2,k} + \boldsymbol{\mu}_{2,k}\boldsymbol{\mu}_{2,k}^T, \tilde{q} \in \mathscr{S}_2\}. && (6.14)
\end{aligned}
$$

We continue to assume that $\mathbf{\Sigma}_1, \mathbf{\Sigma}_{2,k} \succ 0$. We now analyze the two-stage moment robust model with piecewise linear objective and ambiguity sets $\mathscr{P}_1$ and $\mathscr{P}_{2,k}$ defined in (6.13)-(6.14), and focus on the cases (i) $\mathscr{S}_1 = \mathbb{R}^{n_1}$ and $\mathscr{S}_2 = \mathbb{R}^{n_2}$, where $n_1 = \dim(\mathbf{x})$ and $n_2 = \dim(\mathbf{w}_k)$; (ii) $\mathscr{S}_1$ and $\mathscr{S}_2$ are described by some ellipsoids, i.e. $\mathscr{S}_1 = \{\mathbf{p} : (\mathbf{p} - \mathbf{p}_0)^T\mathbf{Q}_1(\mathbf{p}-\mathbf{p}_0) \leq 1\}$, $\mathscr{S}_2 = \{\mathbf{q} : (\mathbf{q}-\mathbf{q}_0)^T\mathbf{Q}_2(\mathbf{q}-\mathbf{q}_0) \leq 1\}$, where $\mathbf{p}_0$, $\mathbf{Q}_1$, $\mathbf{q}_0$ and $\mathbf{Q}_2$ are given and matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ have at least one strictly positive eigenvalue. We provide an equivalent two-stage semidefinite reformulation in the following theorem.

Theorem 8 *Assume $\mathbf{\Sigma}_1, \mathbf{\Sigma}_{2,k} \succ 0$ for $\forall k$. Consider the cases (i) $\mathscr{S}^1 = \mathbb{R}^{n_1}$ and $\mathscr{S}^2 = \mathbb{R}^{n_2}$; or (ii) $\mathscr{S}_1 = \{\mathbf{p} : (\mathbf{p}-\mathbf{p}_0)^T\mathbf{Q}_1(\mathbf{p}-\mathbf{p}_0) \leq 1\}$, $\mathscr{S}_2 = \{\mathbf{q} : (\mathbf{q}-\mathbf{q}_0)^T\mathbf{Q}_2(\mathbf{q}-\mathbf{q}_0) \leq 1\}$, where $\mathbf{p}_0$, $\mathbf{Q}_1$, $\mathbf{q}_0$ and $\mathbf{Q}_2$ are given and matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ have at least one strictly positive eigenvalue. Then, the two-stage moment robust problem* (6.1)-(6.3) *is equivalent to the two-stage semidefinite programming problem:*

$$
\begin{aligned}
&\min_{\mathbf{x},\mathbf{Y}_1,\mathbf{y}_1,y_1^0} \ (\mathbf{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T) \bullet \mathbf{Y}_1 + \boldsymbol{\mu}_1^T\mathbf{y}_1 + y_1^0 + G(\mathbf{x}), && (6.15)\\
&\text{s.t.} \ \begin{pmatrix} \mathbf{Y}_1 & \frac{\mathbf{y}_1 - c_i\mathbf{x}}{2} \\ \frac{(\mathbf{y}_1 - c_i\mathbf{x})^T}{2} & y_1^0 - d_i \end{pmatrix} \succeq \tau_i \mathbf{B}_1, \ \text{for } i = 1,\ldots,M,\\
&\quad \mathbf{x} \in \mathcal{X},
\end{aligned}
$$

where

$$
\begin{aligned}
&G(\mathbf{x}) := \sum_{k=1}^K \pi_k G_k(\mathbf{x}), && (6.16)\\
&G_k(\mathbf{x}) := \min_{\mathbf{w}_k,\mathbf{Y}_{2,k},\mathbf{y}_{2,k},y_{2,k}^0} \ (\mathbf{\Sigma}_{2,k} + \boldsymbol{\mu}_{2,k}\boldsymbol{\mu}_{2,k}^T) \bullet \mathbf{Y}_{2,k} + \boldsymbol{\mu}_{2,k}^T\mathbf{y}_{2,k} + y_{2,k}^0, && (6.17)\\
&\text{s.t.} \ \begin{pmatrix} \mathbf{Y}_{2,k} & \frac{\mathbf{y}_{2,k} - c_i\mathbf{w}_k}{2} \\ \frac{(\mathbf{y}_{2,k} - \mathbf{w}_k)^T}{2} & y_{2,k}^0 - d_i \end{pmatrix} \succeq \tau_i \mathbf{B}_2 \ \text{for } i = 1,\ldots,M,\\
&\quad \mathbf{w}_k \in \mathcal{W}_k(\mathbf{x}),
\end{aligned}
$$

where $\mathbf{B}_1$ and $\mathbf{B}_2$ are defined in Theorem 7.

Proof Given $\mathbf{x} \in \mathcal{X}$, the dual of the inner problem (6.18)

$$
\max_{\mathbb{P}\in\mathscr{P}_1} \mathbb{E}_{\mathbb{P}}[U(\tilde{\mathbf{p}}^T\mathbf{x})] \tag{6.18}
$$

can be written as:

$$
\begin{aligned}
&\min_{\mathbf{Y}_1,\mathbf{y}_1,y_1^0} (\mathbf{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T) \bullet \mathbf{Y}_1 + \boldsymbol{\mu}_1^T\mathbf{y}_1 + y_1^0, && (6.19)\\
&\text{s.t.} \ \mathbf{p}^T\mathbf{Y}_1\mathbf{p} + \mathbf{p}^T\mathbf{y}_1 + y_1^0 \geq U(\mathbf{q}^T\mathbf{x}) \ \ \text{for } \forall \mathbf{q} \in \mathscr{S}_1. && (6.20)
\end{aligned}
$$

Since $U(z) = \max_{i=1,\ldots,M}\{c_i z + d_i\}$, constraint (6.20) is equivalent to:

$$
\mathbf{p}^T\mathbf{Y}_1\mathbf{p} + \mathbf{p}^T\mathbf{y}_1 + y_1^0 \geq c_i\mathbf{q}^T\mathbf{x} + d_i \ \ \text{for } \forall \mathbf{q} \in \mathscr{S}_1, i = 1,\ldots,M. \tag{6.21}
$$

Applying Theorem 5 to (6.21), we know that the inner problem (6.18) is equivalent to the semidefinite programming problem

$$\min_{\mathbf{Y}_1, \mathbf{y}_1, y_1^0} (\boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^T) \bullet \mathbf{Y}_1 + \boldsymbol{\mu}_1^T \mathbf{y}_1 + y_1^0, \tag{6.22}$$

$$\text{s.t.} \begin{pmatrix} \mathbf{Y}_1 & \frac{\mathbf{y}_1 - c_i \mathbf{x}}{2} \\ \frac{(\mathbf{y}_1 - c_i \mathbf{x})^T}{2} & y_1^0 - d_i \end{pmatrix} \succeq \tau_i \mathbf{B}_1, \text{ for } i = 1, \ldots, M. \tag{6.23}$$

Similarly, we can prove that for each given $\mathbf{x} \in \mathcal{X}$, $k = 1, \ldots, M$, $\mathbf{w}_k \in \mathcal{W}_k(\mathbf{x})$, the inner problem

$$\max_{\mathbb{P} \in \mathscr{P}_{2,k}} \mathbb{E}_{\mathbb{P}}[U(\tilde{\mathbf{q}}^T \mathbf{w}_k)] \tag{6.24}$$

is equivalent to the semidefinite programming problem

$$\min_{\mathbf{Y}_{2,k}, \mathbf{y}_{2,k}, y_{2,k}^0} (\boldsymbol{\Sigma}_{2,k} + \boldsymbol{\mu}_{2,k} \boldsymbol{\mu}_{2,k}^T) \bullet \mathbf{Y}_{2,k} + \boldsymbol{\mu}_{2,k}^T \mathbf{y}_{2,k} + y_{2,k}^0, \tag{6.25}$$

$$\text{s.t.} \begin{pmatrix} \mathbf{Y}_{2,k} & \frac{\mathbf{y}_{2,k} - c_i \mathbf{w}_k}{2} \\ \frac{(\mathbf{y}_{2,k} - c_i \mathbf{w}_k)^T}{2} & y_{2,k}^0 - d_i \end{pmatrix} \succeq \tau_i \mathbf{B}_2, \text{ for } i = 1, \ldots, M. \tag{6.26}$$

We can get the desired equivalent two-stage semidefinite programming reformulation (6.10)-(6.12) by combining the equivalent reformulations (6.22)-(6.23) and (6.25)-(6.26) with the outer problem. ∎

6.2 A Two-Stage Portfolio Optimization Problem

In this section we study a two-stage portfolio optimization problem based on model (6.1)-(6.3). The problem is described as follows. An investor needs to plan a portfolio of n assets for two periods. In the first period, the first two moments, $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^T$ are the known information of the probability distribution of the return vector $\mathbf{p} = (p_1, \ldots, p_n)^T$. The probability distribution of the return vector $\mathbf{p}$ can be any probability measure in the ambiguity set $\mathscr{P}_1$ defined in (6.13). The magnitude of the variation between the investment strategy $\mathbf{x}$ and initial strategy $\mathbf{x}_0$ should not excede δ, i.e. $||\mathbf{x} - \mathbf{x}_0|| \leq \delta$, to maintain investment stability and reduce trading costs. The investor reinvests in the n assets at the beginning of the second period. Similarly, the deviation between the investment strategy of this period and the strategy $\mathbf{x}$ of the last period should not exceed δ. The probability distribution of the return vector of the second period depends on some scenario $\boldsymbol{\xi}$ drawn from distribution $\mathbb{D}$. Assume that the sample space of distribution $\mathbb{D}$ consists of K scenarios $\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^K$. Let the probability distribution of the return vector $\mathbf{q} = (q_1, \ldots, q_n)^T$ be in the ambiguity set $\mathscr{P}_{2,k}$ defined in (6.13). We assume that the investor is risk-averse and use a piecewise linear concave utility function as: $u(y) = \min_{i \in \{1,\ldots,M\}} a_i^T y + b_i$. The total utility of the investor is the sum of utilities from both stages. We consider two options for the choices of $\mathscr{S}_1$ and $\mathscr{S}_2$: either $\mathscr{S}_1 = \mathscr{S}_2 = \mathbb{R}^n$ or $\mathscr{S}_1 = \{\mathbf{p} : (\mathbf{p} - \mathbf{p}_0)^T \mathbf{Q}_1 (\mathbf{p} - \mathbf{p}_0) \leq 1\}$, $\mathscr{S}_2 = \{\mathbf{q} : (\mathbf{q} - \mathbf{q}_0)^T \mathbf{Q}_2 (\mathbf{q} - \mathbf{q}_0) \leq 1\}$, where matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ have at least one strictly positive eigenvalue. Therefore, we model this investor's problem as:

$$\min_{\mathbf{x}} \max_{\mathbb{P}\in\mathscr{P}_1} \mathbb{E}_{\mathbb{P}}[\max_i -a_i\tilde{\mathbf{p}}^T\mathbf{x} - b_i] + G(\mathbf{x}), \tag{6.27}$$
$$\text{s.t. } ||\mathbf{x}-\mathbf{x}_0|| \le \delta,$$
$$\mathbf{e}^T\mathbf{x} = 1,\ \mathbf{x}_l \le \mathbf{x} \le \mathbf{x}_u,$$
$$G(\mathbf{x}) := \sum_{k=1}^{K} \pi_k G_k(\mathbf{x}), \tag{6.28}$$
$$G_k(\mathbf{x}) := \min_{\mathbf{w}_k} \max_{\mathbb{P}\in\mathscr{P}_{2,k}} \mathbb{E}_{\mathbb{P}}[\max_i -a_i\tilde{\mathbf{q}}^T\mathbf{w}^{\boldsymbol{\xi}} - b_i], \tag{6.29}$$
$$\text{s.t. } ||\mathbf{x}-\mathbf{x}_0|| \le \delta,$$
$$\mathbf{e}^T\mathbf{w}_k = 1,\ \mathbf{w}_{k,l} \le \mathbf{w}_k \le \mathbf{w}_{k,u},$$

where $\mathbf{e}$ is the n-dimensional vector with 1 in each entry, and $\mathbf{x}_l$, $\mathbf{x}_u$, $\mathbf{w}_{k,l}$, $\mathbf{w}_{k,u}$ are bounds on the investments. A direct application of Theorem 8 results in the following theorem.

Theorem 9 *Assume* $\boldsymbol{\Sigma}_1,\ \boldsymbol{\Sigma}_{2,k} \succ 0$ *for* $k = 1,\ldots,M$. *The two stage portfolio optimization problem* (6.27)-(6.29) *is equivalent to the two-stage stochastic semidefinite programming problem:*

$$\min_{\mathbf{x},\mathbf{Y}_1,\mathbf{y}_1,y_1^0} (\boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T) \bullet \mathbf{Y}_1 + \boldsymbol{\mu}_1^T\mathbf{y}_1 + y_1^0 + G(\mathbf{x}), \tag{6.30}$$
$$\text{s.t. } \begin{pmatrix} \mathbf{Y}_1 & \frac{1}{2}(\mathbf{y}_1 + a_i\mathbf{x}) \\ \frac{1}{2}(\mathbf{y}_1 + a_i\mathbf{x})^T & y_1^0 + b_i \end{pmatrix} \succeq \tau_i\mathbf{B}_1, i = 1,\ldots,M,$$
$$||\mathbf{x}-\mathbf{x}_0|| \le \delta,$$
$$\mathbf{e}^T\mathbf{x} = 1, \mathbf{x}_l \le \mathbf{x} \le \mathbf{x}_u,$$
$$\tau_i \ge 0, \forall i = 1,\ldots,M,$$
$$G(\mathbf{x}) = \sum_{k=1}^{K} \pi_k G_k(\mathbf{x}), \tag{6.31}$$
$$G_k(\mathbf{x}) := \min_{\mathbf{w}_k,\mathbf{Y}_{2,k},\mathbf{y}_{2,k},y_{2,k}^0} (\boldsymbol{\Sigma}_{2,k} + \boldsymbol{\mu}_{2,k}\boldsymbol{\mu}_{2,k}^T) \bullet \mathbf{Y}_{2,k} + \boldsymbol{\mu}_{2,k}^T\mathbf{y}_{2,k} + y_{2,k}^0, \tag{6.32}$$
$$\text{s.t. } \begin{pmatrix} \mathbf{Y}_{2,k} & \frac{1}{2}(\mathbf{y}_{2,k} + a_i\mathbf{w}_k) \\ \frac{1}{2}(\mathbf{y}_{2,k} + a_i\mathbf{w}_{\mathbf{k}})^T & y_{2,i}^0 + b_i \end{pmatrix} \succeq \eta_{i,k}\mathbf{B}_2, i = 1,\ldots,M,$$
$$||\mathbf{w}_k - \mathbf{x}|| \le \delta,$$
$$\mathbf{e}^T\mathbf{w}_k = 1,\ \mathbf{w}_{k,l} \le \mathbf{w}_k \le \mathbf{w}_{k,u},$$
$$\eta_{i,k} \ge 0, \forall i = 1,\ldots,M,$$

where

(1) $\mathbf{B}_1 = \mathbf{B}_2 = 0$ *if* $\mathscr{S} = \mathbb{R}^n$;

(2) $\mathbf{B}_1 = \begin{pmatrix} \mathbf{Q}_1 & -\mathbf{Q}_1\mathbf{p}_0 \\ -\mathbf{p}_0^T\mathbf{Q}_1 & \mathbf{p}_0^T\mathbf{Q}_1\mathbf{p}_0 - 1 \end{pmatrix}$, $\mathbf{B}_2 = \begin{pmatrix} \mathbf{Q}_2 & -\mathbf{Q}_2\mathbf{q}_0 \\ -\mathbf{q}_0^T\mathbf{Q}_2 & \mathbf{q}_0^T\mathbf{Q}_2\mathbf{q}_0 - 1 \end{pmatrix}$, *if* $\mathscr{S}_1 = \{\mathbf{p} : (\mathbf{p} - \mathbf{p}_0)^T\mathbf{Q}_1(\mathbf{p}-\mathbf{p}_0) \le 1\}$, $\mathscr{S}_2 = \{\mathbf{q} : (\mathbf{q}-\mathbf{q}_0)^T\mathbf{Q}_2(\mathbf{q}-\mathbf{q}_0) \le 1\}$.

6.3 Computational Study

In this section the two stage model (6.27)-(6.29) is numerically studied with practical data. We first choose a multivariate AR-GARCH model to forecast the second stage mean vector $\boldsymbol{\mu}_{2,k}$ and covariance matrix $\boldsymbol{\Sigma}_{2,k}$ from the first stage moments $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$. Then we compare the performance of our two stage moment robust model with the other two models, i.e. (1) stochastic programming model and (2) single stage moment robust model. Our empirical results suggest that our two-stage moment robust modeling framework performs better when we have predictive power.

6.3.1 Establishing Parameters of the Two-Stage Portfolio Optimization Model

In the numerical example, the vectors $\mathbf{p}$ and $\mathbf{q}$ in (6.27)-(6.29) are return vectors in period t and $t+1$. The return process $\mathbf{r}_t$ is described by a multivariate AR-GARCH model as follows:

$$\mathbf{r}_t = \phi_0 + \phi_1 \mathbf{r}_{t-1} + \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \sim N(0, \mathbf{Q}_t), \tag{6.33}$$

$$\mathbf{Q}_t = \mathbf{C}\mathbf{C}^T + \mathbf{A}^T \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}^T \mathbf{A}^T + \mathbf{B}^T \mathbf{Q}_{t-1}\mathbf{B}. \tag{6.34}$$

The expected return $\mathbf{r}_t$ is predicted by using a multivariate AR model and the covariance is predicted by a multivariate BEKK GARCH model. The data set is a historical data set of 3 assets over a 7-year horizon (2006-2013), obtained from Yahoo! Finance website. The 3 assets are: AAR Corp., Boeing Corp. and Lockheed Martin. The basic calibration strategy is to use least squares to solve the multivariate AR model to get the residuals $\boldsymbol{\epsilon}$ and then use $\boldsymbol{\epsilon}$ as an input for the BEKK GARCH model [9]. The return model is described as follows. At the beginning of a certain day t, we estimate the expected return $\mathbf{r}_t$ and covariance matrix $\mathbf{Q}_t$ of the current day by using the data of the last 30 days. We then use the AR-GARCH model (6.33)-(6.34) to forecast $\mathbf{Q}_{t+1}$ and then $\mathbf{r}_{t+1}$ follow the distribution $N(\phi_0 + \phi_1 \mathbf{r}_t, \mathbf{Q}_{t+1})$. We start by generating $(\mathbf{r}_t + 1)$ using an n-dimensional Sobol' sequence [4, 6]. We start to generate a set of K n-dimensional Sobol' points $(\mathbf{S}_1, \ldots, \mathbf{S}_K)$. Then we set $\boldsymbol{\epsilon}_i = \mathbf{Q}_{t+1}^{1/2}\Phi^{-1}(S_i)$, $i = 1, \ldots, K$, where Φ is the cumulative normal distribution function. Finally, we generate K samples of $\mathbf{r}_{t+1}$, i.e. $\mathbf{r}_{t+1,1}, \ldots, \mathbf{r}_{t+1,K}$ by using (6.35).

$$\mathbf{r}_{t+1,i} = \phi_0 + \phi_1 \mathbf{r}_t + \boldsymbol{\epsilon}_i,\ i = 1, \ldots, K. \tag{6.35}$$

At the beginning of day t, we optimize the investment strategy $\mathbf{x}$ by considering the forecasts of day $t+1$. We will use the data from $t-750$ to $t-1$ (around 3 year) to calibrate the model (6.33)-(6.34). We re-calibrate the model at the beginning of each day. The two-stage linear conic programming model (6.30)-(6.32) is solved by using SeDuMi [18]. In particular, the parameters are chosen as follows, $\mathscr{S}_1 = \mathscr{S}_2 = \mathbb{R}^n$, $\delta = 1$, $\mathbf{x}_l = \mathbf{w}_{k,l} = -\mathbf{e}$ and $\mathbf{x}_u = \mathbf{w}_{k,u} = \mathbf{e}$ for $\forall k$. Note that the lower bounds $\mathbf{x}_l$ and $\mathbf{w}_{k,u}$ are negative since the investor is allowed to take a short position. We start from 2009 and solve the problem for each day in 2009-2012.

6.3.2 Static Models

We compare our two-stage model (6.27)-(6.29) with the single stage moment robust model which is described as:

$$\begin{aligned}\min_{\mathbf{x}} \max_{\mathbb{P}\in\mathscr{P}_1} & \mathbb{E}_{\mathbb{P}}[U(\mathbf{p}^T\mathbf{x})],\\ \text{s.t. } & ||\mathbf{x}-\mathbf{x}_0|| \le \delta,\\ & \mathbf{e}^T\mathbf{x} = 1,\\ & \mathbf{x}_l \le \mathbf{x} \le \mathbf{x}_u,\end{aligned}$$

where $U(\cdot)$ is the piecewise-linear concave function described in Section 6.1 and $\mathscr{P}_1$ is the probability ambiguity set defined in (6.13).

6.3.3 Evaluating the Significance of the Two-Stage Moment Robust Model

Computational results for the two stage robust model and the static models are shown in Figures 1 and 2 for a three asset and a ten asset problem. These results show that in the case of the three asset problem the returns from the two-stage model are better than those from the static model. However, in the case of the ten asset problems the returns are not significantly different. This is because in the two-case model the AR-GARCH model appears to have greater predictability of returns on this subset of assets than in the case of the ten asset problem. Nevertheless, this example illustrates that the two-stage robust model can out-perform the static robust model when we have future predictive ability in the system.

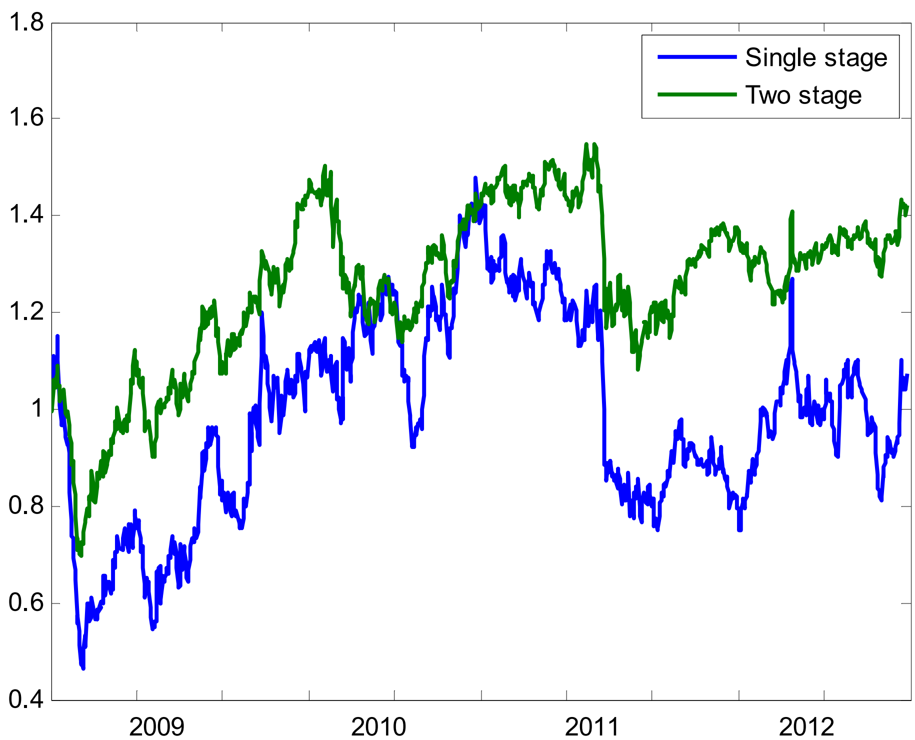

Figure 1: The summary of comparison in 2009-2012 (3 assets, i.e. AAR Corp., Boeing Corp., Lockheed Martin).

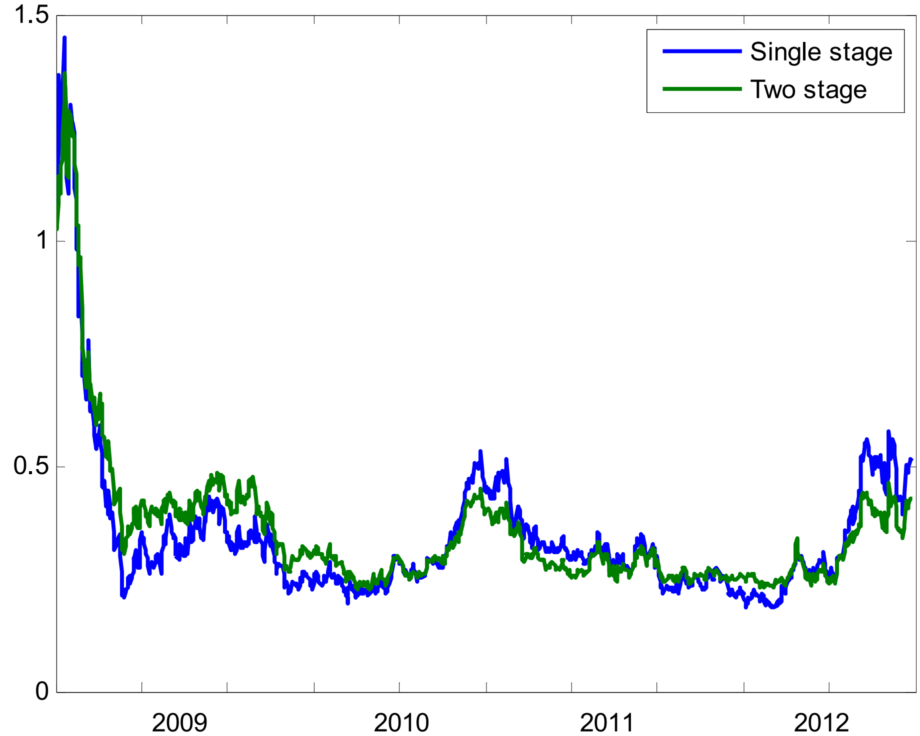

Figure 2: The summary of comparison in 2009-2012 (10 assets, i.e. AAR Corp., Boeing Corp., Lockheed Martin, United Technologies, Intel Corp., Hitachi, Texas Instruments, Dell, Hewlett Packard and IBM Corp).

6.3.4 Algorithmic Performance

We summarize the computational performance of the two-stage model for both 3-asset and 10-asset problems by generating $K = 100; 200; 500; 1000$ samples for the second stage problem. The results are summarized in Table 1. We find that that the average number of IPM iterations do not change with the number of second stage scenarios. We also find that the average runtime increases linearly with the second stage sample size. Both facts implies that our two-stage moment robust model can be solved efficiently by applying the interior point method.

Table 1: Summary of Computational Results.

Sample Size	Avg Num of IPM Iterations		Average Runtime (sec)	
	3-Asset	10-Asset	3-Asset	10-Asset
100	10.7088	15.2821	2.1780	12.9273
200	9.9468	12.8203	3.4871	12.7866
500	9.4116	11.4237	6.8157	33.8398
1000	8.6606	11.0743	13.1113	79.8326

7 Conclusion

In this paper, we propose a two-stage moment robust optimization model. We show that under certain general assumptions, this model can be solved to any ϵ precision in polynomial time. New analysis was required because the second stage problem could only be solved to ϵ precision. The weak version of the polynomial solvablity theorem of Grotschel and Lovasz [7] for convex programs was needed to prove the polynomial solvability. Although the second stage problem has a discrete support, it can be generalized to the continuous support case by the Sample Average Approximation technique, whose convergence is guaranteed (see [15] for details). A two-stage portfolio optimization model with piecewise linear objective is presented and practical data are used to prove the effectiveness and solvability of the two-stage moment robust model.

References

[1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming.* Wiley-Interscience, 3rd edition, 2006.

[2] 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(3):580–602, 2010. Mathematics of Operations Research.

[3] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[4] Paul Bratley and Bennett L. Fox. Implementing sobol's quasirandom sequence generator. *ACM Transactions on Mathematical Software*, 14(1):88–100, 1988.

[5] E. Delage and Y. Ye. Distributionally robust opotimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[6] Paul Glasserman. *Monte Carlo Methods in Financial Engineering.* Springer, 2003.

[7] M. Grotschel, L. Lovasz, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197, 1981.

[8] Jean-Baptiste Hiriart-Urruty and Claude Lemarechal. *Convex Analysis and Minimization Algorithms: Part 2: Advanced Theory and Bundle Methods.* Springer, 1993.

[9] Alexander J. McNeil, Rudiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton University Press, 2005.

[10] S. Mehrotra and M. Gokhan Ozevin. Decomposition-based interior point methods for two-stage stochastic semidefinite programming. *SIAM Journal of Optimization*, 18:206–222, 2007.

[11] S. Mehrotra and M. Gokhan Ozevin. Decomposition-based interior point methods for two-stage stochastic convex quadratic programs with recourse. *Operations Research*, 57(4):964–974, 2009.

[12] S. Mehrotra and M. Gokhan Ozevin. On the implementation of interior point decomposition algorithms for two-stage stochastic conic programs. *SIAM Journal of Optimization*, 19:1846–1880, 2009.

[13] Sanjay Mehrotra and He Zhang. Models and algorithms for distributionally robust least square problems. *Accepted*, 2011.

[14] Georg Pflug and David Wozabal. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442, 2007.

[15] A. Ruszczynski and A. Shapiro, editors. *Stochastic Programming.* Handbooks in Operations Research and Management Science 10. North-Holland, Amsterdam, 2003.

[16] H. Scarf. *Studies in the Mathematical Theory of Inventory and Production*, chapter A min-max solution of an inventory problem, pages 201–209. Stanford University Press, 1958.

[17] A. Shapiro and S. Ahmed. On a class of minimax stochastic programs. *SIAM Journal of Optimization*, 14(4):1237–1249, 2004.

[18] Jos F. Sturm. Using SeDuMi 1.02, a MATLAB* toolbox for optimization over symmetric cones. *Optim. Methods Softw.*, 11-12:625–653, 1999.