Fulfillment-oriented Warehouse Network Design with Demand Information Uncertainty and Its Application in Upfront Warehouse Decision Making

Du Chen

Antai College of Economics and Management, Shanghai Jiao Tong University, China

Abstract

Disaster relief network, if properly designed, can save thousands of lives. However, disaster relief network design is hard due to the unavailability of demand information. In this paper, we study a network design problem for disaster relief purpose when only the support set, first moment and second moment informations are known to decision makers. A two-stage distributionally robust optimization problem is proposed. By utilizing properties of copositive cone (CO), which is widely adapted to study moment-based distributionally robust problems, the proposed model is equivalently reformulated to a CO programming model under modest assumptions. A case study shows that, in terms of total cost and second-type service level, our CO model with limited demand information is comparable to a two-stage stochastic model with full demand information. It also significantly outperforms a robust Ψ-expander model with only support set information.

Keywords: Distribution network design, Distributionally robust optimization, Copositive cone

1. Introduction

2. Literature Review

2.1. Network Design for Disaster Management

A network is *balanced* if there are an equal number of potential warehosues and demand areas, and the system is *asymmetric* if not all demand areas have i.i.d demand.

2.2. Distributionally Robust Optimization

2.3. Copositive Programming

3. Fulfillment-Oriented Network Design Problem

Consider a fulfillment-oriented warehouse network design problem with a set of potential supply nodes denoted by $\mathcal{W} = \{1, 2, \ldots, m\}$ and a set of demand nodes denoted by $\mathcal{R} = \{1, 2, \ldots, n\}$. The road links is

Email address: `chendu2017@sjtu.edu.cn` (Du Chen)

denoted by $\mathcal{E}$, and its cardinality is $|\mathcal{E}| = r$. We denote $\Gamma(i)$ as the set of adjacent nodes for location i. Warehouse $i \in \mathcal{W}$ is able to deliver products to all adjacent demand nodes $\Gamma(i)$ through roads. Correspondingly, a demand node $j \in \mathcal{R}$ can receive products from its adjacent warehouses $\Gamma(j)$. Suppose the decision maker has only limited information about the demand, i.e. only the first-moment, second-moment, and support set are known. Facing the uncertainty of future demand, the decision maker should select a subset of $\mathcal{W}$ to build large-enough warehouses, and proactively deploy products accordingly, so that after demand realized, the decision maker could allocate pre-depolyed products to satisfy as much demand as possible (or equivalently as less unmet demand as possible). Each built warehouse is associated with a fixed setup cost $f_i, i \in [m]$, and each unit prepositioning product causes a holding cost $h_i, i \in [m]$. Since we focus on fulfillment-oriented circumstance, each unmet demand will trigger one unit penalty cost. We want to minimize the total cost, including first-stage setup cost and the worst expected second-stage penalty cost. We temporarily assume there is only one kind of products, and we will consider multi-items situation in Section 5.1.

3.1. Model

The fulfillment-oriented network design (FOND) problem under demand distribuiton uncertainty, given the ambiguity set is characterized by first- and second-moment, can be formulated as:

$$\begin{aligned} (FOND) \quad \min_{\boldsymbol{I},\boldsymbol{Z}} \quad & \kappa(\boldsymbol{f^T Z} + \boldsymbol{h^T I}) + \sup_{\mathbb{P} \in \mathscr{F}(\boldsymbol{\mu},\boldsymbol{\Sigma})} \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{I},\boldsymbol{Z},\tilde{\boldsymbol{d}})\right] \\ \text{s.t.} \quad & \boldsymbol{I} \le M\boldsymbol{Z}, \\ & \boldsymbol{I} \in \mathbb{R}_+^m, \boldsymbol{Z} \in \{0,1\}^m \end{aligned} \tag{1}$$

where the distribution ambiguity set is:

$$\mathscr{F}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \left\{ \mathbb{P} \in \mathcal{P}\left(\mathbb{R}_+^n\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{d}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{d}}] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{d}}\tilde{\boldsymbol{d}}^T] = \boldsymbol{\Sigma} \\ \mathbb{P}[\tilde{\boldsymbol{d}} \in \mathbb{R}_+^n] = 1 \end{array} \right\}$$

In Model (1), $\boldsymbol{f}$ is the setup cost of building a new warehouse, $\boldsymbol{h}$ is the holding cost for each unit product, and $\kappa \ge 0$ is a risk attitude parameter balacing costs of two stages. $\boldsymbol{I}$ is the decision variable, meaning the quantity of relief materials that should be deployed at each location. Another decision variable, $\boldsymbol{Z}$, represents whether a new warehouse should be built. The first constraint guarantees that materials can be stored at locations where a warehouse is establied. Remaining constraints are standard nonegative and binary constraints. $g(\boldsymbol{I},\boldsymbol{Z},\tilde{\boldsymbol{d}})$ is the number of unmet demand obtained by solving an classical allocation problem after demand $\tilde{\boldsymbol{d}}$ is realized. And the allocation problem can be described as:

$$g(\boldsymbol{I},\boldsymbol{Z},\tilde{\boldsymbol{d}}) = \min_{\boldsymbol{x} \in \Omega(\boldsymbol{I},\boldsymbol{Z})} \sum_{j \in [n]} \tilde{d}_j - \sum_{(i,j) \in \mathcal{E}(\boldsymbol{Z})} x_{ij} \tag{2}$$

where $\Omega(\boldsymbol{I}, \boldsymbol{Z})$ is the feasible region characterized by

$$\Omega(\boldsymbol{I}, \boldsymbol{Z}) := \{\boldsymbol{x} \in \mathbb{R}_+^{|\mathcal{E}(\boldsymbol{Z})|} | \sum_{i:(i,j)\in\mathcal{E}(\boldsymbol{Z})} x_{ij} \leq \tilde{d}_j, \forall j \in [n]; \sum_{j:(i,j)\in\mathcal{E}(\boldsymbol{Z})} x_{ij} \leq I_i, \forall i \in [m]\}$$

$\mathcal{E}(\boldsymbol{Z})$ is the accessible roads based on decision $\boldsymbol{Z}$.

It is clear to see, Model (2) depends both on $\boldsymbol{I}$ and $\boldsymbol{Z}$, and accessible network is varied under different $\boldsymbol{Z}$, which hampers us from solving Model (1). Fortunately, in fact, (2) could get rid of the dependency on $\boldsymbol{Z}$ at the cost of lifting the dimension of feasible region from $|\mathcal{E}(\boldsymbol{Z})|$ to r.

Proposition 1. *Given a feasible solution* $(\boldsymbol{I}, \boldsymbol{Z})$ *to* (1)*, for any demand realization* $\tilde{\boldsymbol{d}} \in \mathbb{R}_+^n$*, the value of* (2) *is equal to the objective value of the following problem.*

$$g(\boldsymbol{I}, \tilde{\boldsymbol{d}}) = \min_{\boldsymbol{x}\in\Omega(\boldsymbol{I})} \sum_{j\in[n]} \tilde{d}_j - \sum_{(i,j)\in\mathcal{E}} x_{ij}$$

where

$$\Omega(\boldsymbol{I}) := \{\boldsymbol{x} \in \mathbb{R}_+^r | \sum_{i:(i,j)\in\mathcal{E}} x_{ij} \leq \tilde{d}_j, \forall j \in [n]; \sum_{j:(i,j)\in\mathcal{E}} x_{ij} \leq I_i, \forall i \in [m]\}$$

Proof. See Appendix A.1 □

According to Proposition 1, instead of considering both $(\boldsymbol{I}, \boldsymbol{Z})$, we only need to take $\boldsymbol{I}$ into consideration. And getting rid of $\boldsymbol{Z}$ does not harm the optimality, however, at the cost of an increasing in the dimension of recourse decision variables $\boldsymbol{x}$. Surprisingly, the increased variables not only make further reformulation easier, but also provides more insights on selection locations for warehouses, which actually accelarates the solving process. We will furthur explain this in Section 3.3.

From now on, we replace $g(\boldsymbol{I}, \boldsymbol{Z}, \tilde{\boldsymbol{d}})$ with $g(\boldsymbol{I}, \tilde{\boldsymbol{d}})$. With the more tractable recourse problem, we next equivalently reformulate Model (1) into a mixed integer copositive programming problem.

3.2. Reformulation

Obviously, the difficulty of Model (1) comes from evaluating the worst expected second-stage unmet demand. Denote the worst-case expected second-stage problem as

$$L(\boldsymbol{I}) := \sup_{\mathbb{P}\in\mathscr{F}(\boldsymbol{\mu},\boldsymbol{\Sigma})} \mathbb{E}_{\mathbb{P}} \left[g(\boldsymbol{I}, \tilde{\boldsymbol{d}})\right] \tag{3}$$

which still only depends on $\boldsymbol{I}$ since the expectation operator does not invalidate Proposition 1. We first analyze the inner allocation problem, which has many properties help up simply the whole problem. Equivalently, $g(\boldsymbol{I}, \tilde{\boldsymbol{d}})$ can be rewritten as

$$
\begin{aligned}
g(\boldsymbol{I}, \tilde{\boldsymbol{d}}) = & \min_{\boldsymbol{x} \in \Omega(\boldsymbol{I})} \sum_{j \in [n]} \tilde{d}_j - \sum_{(i,j) \in \mathcal{E}} x_{ij} \\
= & \sum_{j \in [n]} \tilde{d}_j - \max_{\boldsymbol{x} \in \Omega(\boldsymbol{I})} \sum_{(i,j) \in \mathcal{E}} x_{ij} \\
= & \sum_{j \in [n]} \tilde{d}_j - \min_{(\boldsymbol{u},\boldsymbol{v}) \in \mathcal{L}} \tilde{\boldsymbol{d}}^T \boldsymbol{u} + \boldsymbol{I}^T \boldsymbol{v} \\
= & - \min_{(\boldsymbol{u},\boldsymbol{v}) \in \mathcal{L}} \tilde{\boldsymbol{d}}^T (\boldsymbol{u} - \mathbf{1}) + \boldsymbol{I}^T \boldsymbol{v} \\
= & \max_{(\boldsymbol{u},\boldsymbol{v}) \in \mathcal{L}} \tilde{\boldsymbol{d}}^T (\mathbf{1} - \boldsymbol{u}) - \boldsymbol{I}^T \boldsymbol{v}
\end{aligned}
$$

The first equation is exactly the minimizing unmet demand problem after $\tilde{\boldsymbol{d}}$ is realized. The second equation holds by simply exchanging minimizing operator and negative symble. The third equation strictly follows the classical min-cut max-flow theorem, where the feasible region is

$$\mathcal{L} := \{(\boldsymbol{u}, \boldsymbol{v}) \in \{0,1\}^{n+m} | u_j + v_i \geq 1, \forall (i,j) \in \mathcal{E}\}$$

The fourth equation holds by combining realized demand vector $\tilde{\boldsymbol{d}}$, where $\mathbf{1}$ is a vector with all elements equal to 1. And the last equation also comes from exchanging maxmizing operator and the negative symble. One notable reformulation trick here is that, at the third equation, directly taking expectation on $\tilde{d}_j$ is doable if the true demand mean is captured by the outer ambiguity set, which temporarily is the case we consider. However, We will further relax the ambiguity set to a more general case. And the reformulation we adopt here would be more compatible with further study. Additionally, from the perspective of stochastic programming, the feasible region of $\mathcal{L}$ is complete recourse, so that the region is always nonempty.

Due to the max-flow theorem, the feasion region $\mathcal{L}$ can be relaxed to a convex polyhedron without losing optimality by relaxing binary variables to continuous variables. The relaxed region is

$$\mathcal{L}_{lp} := \{(\boldsymbol{u}, \boldsymbol{v}) \in [0,1]^{n+m} | u_j + v_i \geq 1, \forall (i,j) \in \mathcal{E}\}$$

For further analytics, we modify $\mathcal{L}$ by replacing $\mathbf{1} - \boldsymbol{u}$ with decision variables $\hat{\boldsymbol{u}} = \mathbf{1} - \boldsymbol{u}$, and define the new region with $\hat{\boldsymbol{u}}$ as

$$\hat{\mathcal{L}} := \{(\hat{\boldsymbol{u}}, \boldsymbol{v}) \in \{0,1\}^{n+m} | v_i \geq \hat{u}_j, \forall (i,j) \in \mathcal{E}\}$$

Moreover, the linear relaxation of the feasible region is

$$\hat{\mathcal{L}}_{lp} := \left\{ (\hat{\boldsymbol{u}}, \boldsymbol{v}) \in \mathbb{R}_+^{n+m} \;\middle|\; \begin{array}{l} v_i \geq \hat{u}_j, \forall (i,j) \in \mathcal{E} \\ \hat{\boldsymbol{u}} \leq \mathbf{1} \\ \boldsymbol{v} \leq \mathbf{1} \end{array} \right\}$$

Following the classical min-cut max-flow theorem, the optimal solution of $g(\boldsymbol{I}, \tilde{\boldsymbol{d}})$ is the same no matter we optimize in region $\hat{\mathcal{L}}$ or $\hat{\mathcal{L}}_{lp}$. Therefore, $g(\boldsymbol{I}, \tilde{\boldsymbol{d}})$ has the same value, as well as the same optimal solution, among the following problems:

$$g(\boldsymbol{I}, \tilde{\boldsymbol{d}}) = \max_{(\hat{\boldsymbol{u}}, \boldsymbol{v}) \in \hat{\mathcal{L}}} \tilde{\boldsymbol{d}}^T \hat{\boldsymbol{u}} - \boldsymbol{I}^T \boldsymbol{v} = \max_{(\hat{\boldsymbol{u}}, \boldsymbol{v}) \in \hat{\mathcal{L}}_{lp} \cap \{0,1\}^{n+m}} \tilde{\boldsymbol{d}}^T \hat{\boldsymbol{u}} - \boldsymbol{I}^T \boldsymbol{v} = \max_{(\hat{\boldsymbol{u}}, \boldsymbol{v}) \in \hat{\mathcal{L}}_{lp}} \tilde{\boldsymbol{d}}^T \hat{\boldsymbol{u}} - \boldsymbol{I}^T \boldsymbol{v} \tag{4}$$

Unfortunately, though we have relaxed the inner allocation problem from an integer programming to a linear programming with a fixed feasible region $\hat{\mathcal{L}}_{lp}$, the difficuly of evaluating (3) does not reduced. Actually, Proposition 2 shows that problem (3), with feasible region $\hat{\mathcal{L}}_{lp}$, is still a NP-hard problem.

Proposition 2. *Given the feasible region to* (4) *is* $\hat{\mathcal{L}}_{lp}$, *calculating the value of problem* (3) *is still a NP-hard problem.*

Proof. See Appendix A.2 □

Through adding slackness variables $(\boldsymbol{s}^\dagger, \hat{\boldsymbol{u}}^\dagger, \boldsymbol{v}^\dagger) \in \mathbb{R}_+^{r+n+m}$, we can modify $\hat{\mathcal{L}}_{lp}$ to a system only consist of equalities:

$$\hat{\mathcal{L}}_{lp}^\dagger := \left\{ (\hat{\boldsymbol{u}}, \boldsymbol{v}, \boldsymbol{s}, \hat{\boldsymbol{u}}^\dagger, \boldsymbol{v}^\dagger) \in \mathbb{R}_+^{2n+2m+r} \;\middle|\; \begin{array}{l} \hat{u}_j - v_i + s_{ij}^\dagger = 0, \forall (i,j) \in \mathcal{E} \\ \hat{\boldsymbol{u}} + \hat{\boldsymbol{u}}^\dagger = \boldsymbol{1} \\ \boldsymbol{v} + \boldsymbol{v}^\dagger = \boldsymbol{1} \end{array} \right\}$$

Therefore, the expected second-stage penalty value under the worst distribution can be equivalently reformulated as:

$$L(\boldsymbol{I}) = \sup_{\mathbb{P} \in \mathscr{F}(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \mathbb{E}_{\mathbb{P}} \left[g(\boldsymbol{I}, \tilde{\boldsymbol{d}}) \right] = \sup_{\mathbb{P} \in \mathscr{F}(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \mathbb{E}_{\mathbb{P}} \left[\max_{(\tilde{\boldsymbol{u}}, \boldsymbol{v}, \boldsymbol{s}, \hat{\boldsymbol{u}}^\dagger, \boldsymbol{v}^\dagger) \in \hat{\mathcal{L}}_{lp}^\dagger \cap \{0,1\}^{n+m}} \tilde{\boldsymbol{d}}^T \hat{\boldsymbol{u}} - \boldsymbol{I}^T \boldsymbol{v} \right]$$

Please note that uncertainties of the inner problem only exist in the objective function. For simplicity of notations, we denote, with a little bit abuse of notations, the decision variables by $\boldsymbol{x} := (\hat{\boldsymbol{u}}, \boldsymbol{v}, \boldsymbol{s}, \hat{\boldsymbol{u}}^\dagger, \boldsymbol{v}^\dagger) \in \hat{\mathcal{L}}_{lp}^\dagger \subset \mathbb{R}_+^N$, where $N = 2m + 2n + r$. The constraints of $\hat{\mathcal{L}}_{lp}^\dagger$ can also be written in a more general form as $\left\{ \boldsymbol{x} \geq \boldsymbol{0} \mid \boldsymbol{a}_i^T \boldsymbol{x} = b_i, \forall i \in [M] \right\}$, where $M = r + n + m$. We illustrate how we write these vectors in general form by an example.

Example 1. Let us consider a simple bipartite graph as show in Fig 1. There are two warehouses and two points of demand. The inventory level of each warehouse is 3 and 5 respectively, and the realized demand

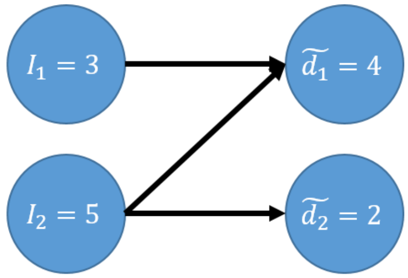

Figure 1: Example: 2×2 Balanced System

for each points are 4 and 2. According to previously defined general form, the equation system should be

$$\begin{bmatrix} \boldsymbol{a}_1^T \\ \boldsymbol{a}_2^T \\ \boldsymbol{a}_3^T \\ \boldsymbol{a}_4^T \\ \boldsymbol{a}_5^T \\ \boldsymbol{a}_6^T \\ \boldsymbol{a}_7^T \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ v_1 \\ v_2 \\ s_{11} \\ s_{21} \\ s_{22} \\ \hat{u}_1^\dagger \\ \hat{u}_2^\dagger \\ v_1^\dagger \\ v_2^\dagger \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ v_1 \\ v_2 \\ s_{11} \\ s_{21} \\ s_{22} \\ \hat{u}_1^\dagger \\ \hat{u}_2^\dagger \\ v_1^\dagger \\ v_2^\dagger \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Before we reformulate $L(\boldsymbol{I})$ to a conic program, we first define three new variables.

- $\boldsymbol{p} := \mathbb{E}_{\mathbb{P}}[\boldsymbol{x}(\tilde{\boldsymbol{d}})] \in \mathbb{R}_+^N$
- $\boldsymbol{Y} := \mathbb{E}_{\mathbb{P}}[\boldsymbol{x}(\tilde{\boldsymbol{d}})\tilde{\boldsymbol{d}}^T] \in \mathbb{R}_+^{N \times n}$
- $\boldsymbol{X} := \mathbb{E}_{\mathbb{P}}[\boldsymbol{x}(\tilde{\boldsymbol{d}})\boldsymbol{x}(\tilde{\boldsymbol{d}})^T] \in \mathbb{R}_+^{N \times N}$

According to Natarajan et al. (2011), we can reformulate $L(\boldsymbol{I})$ to a conic program. We formally propose this reformulation in Proposition 3. Although Natarajan et al. (2011), Kong et al. (2013), Yan et al. (2018), and Kong et al. (2020) utilize the same technique to prove the equivalence, we find that for the *reduced form*, where the number of random coefficient in objective function is less than that of decision variables, a full proof is still missing in literatures. We provide the detailed proof in Appendix A.3.

Proposition 3. *$L(\boldsymbol{I})$ can be equivalently reformulated to a completely positive program $L_{CP}(\boldsymbol{I})$ as follows:*

$$
\begin{aligned}
L(\boldsymbol{I}) = \quad L_{CP}(\boldsymbol{I}) = \quad & \max_{\boldsymbol{p},\boldsymbol{Y},\boldsymbol{X}} \begin{bmatrix} \boldsymbol{e}^{[n]} \\ \boldsymbol{0}_{(N-n)\times n} \end{bmatrix} \bullet \boldsymbol{Y} + \begin{bmatrix} \boldsymbol{0}_{n\times 1} \\ -\boldsymbol{I}_{m\times 1} \\ \boldsymbol{0}_{M\times 1} \end{bmatrix}^T \boldsymbol{p} \\
s.t. \quad & \boldsymbol{a}_i^T \boldsymbol{p} = b_i, \forall i \in [M] && (5) \\
& (\boldsymbol{a}_i \boldsymbol{a}_i^T) \bullet \boldsymbol{X} = b_i^2, \forall i \in [M] && (6) \\
& \boldsymbol{e}_{(j,j)} \bullet \boldsymbol{X} - \boldsymbol{e}_{(j)}^T \boldsymbol{p} = 0, \forall j \in [N] && (7) \\
& \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{p}^T \\ \boldsymbol{\mu} & \boldsymbol{\Sigma} & \boldsymbol{Y}^T \\ \boldsymbol{p} & \boldsymbol{Y} & \boldsymbol{X} \end{pmatrix} \succeq_{cp} \boldsymbol{0} && (8)
\end{aligned}
$$

We use superscribe $^{[k]}$ to define a k-by-k matrix variable. $\boldsymbol{e}^{[n]}$ is a n-dimensional identity matrix. We further use $\boldsymbol{e}_{(j,j)}$ to represent a matrix with (j,j)-position element equal to 1. And $\boldsymbol{e}_{(j)}$ is the unit vector with j-th element equal to 1. $\succeq_{cp}$ means that the matrix should belong to a completely positive cone. Except further explanation, $\boldsymbol{0}$ should be a zero vector or zero matrix with proper shape. Here, we summarize the main idea of proof as follows. By construction, the optimal solution to $L(\boldsymbol{I})$ is a feasible solution to $L_{CP}(\boldsymbol{I})$. Therefore, $L(\boldsymbol{I}) \leq L_{CP}(\boldsymbol{I})$. On the other side, by decomposing the optimal solution of $L_{CP}(\boldsymbol{I})$, we can construct a sequence of random vectors whose limit satisfies the moment condition, as well as a sequence of corresponding feasible solutions. The limit of the set of feasible solutions achieves the objective value of $L(\boldsymbol{I})$ from below, i.e. $L(\boldsymbol{I}) \geq L_{CP}(\boldsymbol{I})$. Therefore, the equivalence is proved.

Taking Duality on model 5 gives the following results

$$
\begin{aligned}
L_{CO}(\boldsymbol{I}) = \quad & \min_{\alpha_i,\beta_i,\theta_j,\tau,\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi},\boldsymbol{\eta},\boldsymbol{w}} \sum_{i\in[M]} b_i\alpha_i + b_i^2\beta_i + \tau + 2\boldsymbol{\mu}^T\boldsymbol{\xi} + \boldsymbol{\Sigma}\bullet\boldsymbol{\eta} \\
s.t. \quad & \sum_{i\in[M]} \boldsymbol{a}_i\alpha_i - \sum_{j\in[N]} \boldsymbol{e}_{(j)}\theta_j - 2\boldsymbol{\varphi} = \begin{bmatrix} \boldsymbol{0}_{n\times 1} \\ -\boldsymbol{I}_{m\times 1} \\ \boldsymbol{0}_{M\times 1} \end{bmatrix} && (9) \\
& \sum_{i\in[M]} \boldsymbol{a}_i\boldsymbol{a}_i^T\beta_i + \sum_{j\in[N]} \boldsymbol{e}_{(j,j)}\theta_j - \boldsymbol{w} = \boldsymbol{0} && (10) \\
& -2\boldsymbol{\psi} = \begin{bmatrix} \boldsymbol{e}^{[n]} \\ \boldsymbol{0} \end{bmatrix} && (11) \\
& \begin{pmatrix} \tau & \boldsymbol{\xi}^T & \boldsymbol{\varphi}^T \\ \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{\psi}^T \\ \boldsymbol{\varphi} & \boldsymbol{\psi} & \boldsymbol{w} \end{pmatrix} \succeq_{co} \boldsymbol{0} && (12)
\end{aligned}
$$

α_i, β_i, and θ_j is the dual variable for constraint (5), (6), and (7) respectively. Variable $\begin{pmatrix} \tau & \boldsymbol{\xi}^T & \boldsymbol{\varphi}^T \\ \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{\psi}^T \\ \boldsymbol{\varphi} & \boldsymbol{\psi} & \boldsymbol{w} \end{pmatrix}$ is the dual variables for completely positive cone constraint (8). By weak duality theorem, $L_{CO}(\boldsymbol{I}) \geq L_{CP}(\boldsymbol{I})$. Surprisingly, $L_{CO}(\boldsymbol{I})$ is equivalent to $L_{CP}(\boldsymbol{I})$

Proposition 4. *$L_{CO}(\boldsymbol{I})$ is equivalent to $L_{CP}(\boldsymbol{I})$*

Proof. See Appendix A.4 □

We further combine $L_{CO}(\boldsymbol{I})$ with first-stage problem and replace a part of decision variables in (12) with equations (9), (10), and (11). Now, we obtain the Mixed-Integer Copositive Cone Problem:

$$(CO) L_{CO} = \min_{\boldsymbol{I},\boldsymbol{Z},\alpha_i,\beta_i,\theta_j,\tau,\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi},\boldsymbol{\eta},\boldsymbol{w}} \boldsymbol{f}^T\boldsymbol{I} + \boldsymbol{h}^T\boldsymbol{Z} + \sum_{i\in[M]} b_i\alpha_i + b_i^2\beta_i + \tau + 2\boldsymbol{\mu}^T\boldsymbol{\xi} + \boldsymbol{\Sigma} \bullet \boldsymbol{\eta}$$

$$s.t. \quad \boldsymbol{I} \leq M\boldsymbol{Z}$$

$$\boldsymbol{I} \geq \boldsymbol{0}, \boldsymbol{Z} \in \{0,1\}^m$$

$$\begin{pmatrix} \tau & \boldsymbol{\xi}^T & \frac{1}{2}\left(\sum_{i\in[M]} \boldsymbol{a}_i\alpha_i - \boldsymbol{\theta}\right)^T \\ \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{0} \\ \frac{1}{2}\left(\sum_{i\in[M]} \boldsymbol{a}_i\alpha_i - \boldsymbol{\theta}\right) & \boldsymbol{0} & \sum_{i\in[M]} \boldsymbol{a}_i\boldsymbol{a}_i^T\beta_i + \boldsymbol{Diag(\theta)} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I} \\ \boldsymbol{0} \end{bmatrix}^T \\ \boldsymbol{0} & \boldsymbol{0} & \begin{bmatrix} -\boldsymbol{e}^{[n]} \\ \boldsymbol{0} \end{bmatrix}^T \\ \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I} \\ \boldsymbol{0} \end{bmatrix} & \begin{bmatrix} -\boldsymbol{e}^{[n]} \\ \boldsymbol{0} \end{bmatrix} & \boldsymbol{0} \end{pmatrix} \succeq_{co} \boldsymbol{0}$$

$\boldsymbol{Diag(\theta)}$ represents a diagnal matrix with i-th diagnal element equal to θ_i. The dimension of vector and matrix strictly follow the same dimension in model 5.

According to Yan et al. (2018) and Dickinson (2010), $L_{CO}(\boldsymbol{I}) = L_{CP}(\boldsymbol{I})$

Proof. according to Yan et al. (2018), if the non-slackness variables (we say *reduced solutions*) of the inner allocation problem has strictly feasible solution in $\mathcal{L}_{lp}^{\dagger}$, and a set of reduced feasible solution can span to $\mathbb{R}^{m+n}$, and has solution $(\hat{\boldsymbol{u}}, \boldsymbol{v}) > 0$, as well as the moment info matrix is in CP cone; then COCone problem = CPCone problem □

3.3. Speed Up Branch&Bound

4. Numerical Study

In this section, we conduct numerous experiments to examine the advantages and disadvantages of the proposed two-stage distributionally robust model, as well as the equivalent conic reformulation. First, We run our models in hundreads of randomly generated balanced but asymmetric networks with predetermined parameters. Second, we conduct sensitivity analysis on three dimensions, including correlation parameter ρ, coefficient of variantion c_v, and risk attitude parameter κ. Finally, we extend our experiments to more general network structures, and re-run all experiments in the first two steps, trying to obtain a comprehensive understanding of proposed models in a more general setting. Throughout this section, a stochastic model and a mean-variance model are introduced as benchmarks, which will be clearly illustrated in the next subsection.

4.1. Experiment Setting

In this subsection, we first introduce a two-stage stochastic model and a mean-variance distributionally robust model as benchmarks. Then, a detailed explanation is given to describe how we generate synthetic data for simulation experiments. Finally, we propose some measurements of interest.

4.1.1. benchmark

We introdcut a two-stage stochastic model and a mean-variance distributionally robust model as benchmarks. The former model fully exploit all demand samples while the later one only involves two statistics, mean and variance, which does not consider correlations between nodes. Since our model incorporte first-moment and cross second-moment information, the information we have is slightly more than the mean-variance model has, but less than the stochastic model has. Hence, we have the chance to evaluate the value of demand information in our problem, to some extent.

SAA with Empirical Distribution

$$
\begin{aligned}
(SAA) \quad \min_{\boldsymbol{I},\boldsymbol{Z}} \quad & \boldsymbol{f^T Z} + \boldsymbol{h^T I} + \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{I}, \tilde{\boldsymbol{d}})\right] \\
\text{s.t.} \quad & \boldsymbol{I} \leq M\boldsymbol{Z}, \\
& \boldsymbol{I} \in \mathbb{R}_+^m, \boldsymbol{Z} \in \{0,1\}^m
\end{aligned}
$$

Mean-Variance

$$
\begin{aligned}
(MV) \quad \min_{\boldsymbol{I},\boldsymbol{Z}} \quad & \boldsymbol{f^T Z} + \boldsymbol{h^T I} + \sup_{\mathbb{P}\in\mathscr{F}(\boldsymbol{\mu},\boldsymbol{\Sigma})} \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{I},\tilde{\boldsymbol{d}})\right] \\
\text{s.t.} \quad & \boldsymbol{I} \leq M\boldsymbol{Z}, \\
& \boldsymbol{I} \in \mathbb{R}_+^m, \boldsymbol{Z} \in \{0,1\}^m
\end{aligned}
$$

4.1.2. Data Generation

Suppose we consider a network with m potential warehouses and n demand nodes. We randomly generate dozens of or hundreds of networks, and each network is defined by a six-elements tuple $(\mathcal{W}, \mathcal{R}, \mathcal{A}, \boldsymbol{f}, \boldsymbol{h}, \boldsymbol{\mu})$. $\mathcal{W}$ and $\mathcal{R}$ are the set of potential warehouse nodes and demand nodes. Temporarily, we assume $|\mathcal{W}| = |\mathcal{R}| = 8$. $\mathcal{A}$ is set of links. According to Ni et al. (2018), the average node degree of a typical road network is about 2.4. This is also align with industrial practices in manufacturing, disaster management, and school bus routing design (Jordan and Graves (1995), Mete and Zabinsky (2010), and Bertsimas et al. (2019a)). Therefore, we require $|\mathcal{A}| = 1.2(m+n)$. Also, the following two constraints on the generated network should be satisfied: (1) at least one road linked to each supply node (2) at least one road linked to each demand node. These two conditions help avoid the appearance of isolated nodes. $\boldsymbol{f}$ is the fixed setup cost randomly draw from a uniform distribution U[50, 100]. $\boldsymbol{h}$ is the holding cost drawn from U[0.1, 0.2]. And $\boldsymbol{\mu}$ is the first moment, i.e. the mean value, of demand, drawn from U[400, 600]. Initially, we set the correlation coefficient $\rho = 0.3$, coefficient of variation $c_v = 0.3$, and risk attitude coefficient $\kappa = 1.0$, all of which will be further modified for sensitivity analysis. For simplicity, we assume the correlation parameter ρ and coefficient of variation c_v are applied to all demand node pairs or demand node, which means the covariance of any two demand nodes (i, j) is $\mu_i \mu_j c_v^2 \rho,\ \forall i, j \in \mathcal{R}$. We further assume the underlying demand distirbution follows a multinormial gaussian distribution, whose domain is defined as $\mathbb{R}_+^n$, the mean value is $\boldsymbol{\mu}$, and the covariance matrix is $\boldsymbol{\Sigma}_{\rho,c_v}$ as defined above. We use the subscripts ${}_{\rho,c_v}$ to emphasize that the covariance matrix depends on ρ and c_v. For model **??**, we suppose the first-moment $\boldsymbol{\mu}$ and second-moment matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\rho,c_v} + \boldsymbol{\mu}\boldsymbol{\mu}^T$ are known to decision makers. For model 13, only mean and variance of demand are known. And for model 13, we assume there are 30 historical samples available. After obtaining optimal solutions through solving three models, we randomly generate 2000 demand realizations to conduct the second-stage simulations. 1000 in-sample demand realizations are drawn from the true distribution $\boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\rho,c_v})$, while another 1000 out-sample realizations are drawn from drawn from what distributions?. We seperately analyze results on in-sample and out-sample realizations, and for each network, we regard the average (or worst-cast) result on 1000 second-stage simulations as the corresponding graph's performance.

4.2. (8, 8) Network

4.2.1. (ρ, c_v, κ)*=(0.3, 0.3, 1.0)*

Following the initial setting as depticted in the above subsection, we first restrict ourself to the (8, 8) structure for a closer look at the model's performance. We randomly generate 200 networks with $(\rho, c_v, \kappa) = (0.3, 0.3, 1.0)$. Except furthur explanations, all analysis in this subsection rely on the 200 networks: for each network, we construct a measurement on results of simulation, for example the average unmet demand, and then take average on 200 networks.

First of all, we compare the first-stage deployments and characteristics of designed networks. Table 1 summarized the first-stage decisions and results of networks. The first column named by *total inv.* represents

Table 1: Summary on Designed Networks

	total inv.	# of w.h.	# of roads	w.h. degree	d.n. degree
SAA	4903.43	4.06	11.57	2.91	1.45
	(300.33)	(0.78)	(1.57)	(0.45)	(0.20)
CO	4944.17	3.79	10.96	2.96	1.37
	(201.06)	(0.72)	(1.52)	(0.48)	(0.19)
MV	4638.28	3.77	11.15	3.02	1.39
	(194.14)	(0.72)	(1.44)	(0.48)	(0.18)

the average total inventory of 200 networks. The number in the parenthesis shows the standard deviation. It is clear to see that Model (CO) result in the highest inventory level while Model (MV) has the lowest inventory level. The difference comes from whether the strong correlation of demand is taken into account. The second column, *# of w.h.*, reveals the number of established warehouses on average. Model (SAA) selects almost half of potential locations to build warehouses. The third columns is the amount of roads in designed networks, which implies that Model (CO) designs the most sparse network. The last two columns exhibit supply nodes' and demand nodes' degree repsectively.

We next check the cost performances, including first-stage deployment cost and second-stage penalty due to unmet demand, which are the most concerned. Table 2 shows the average costs on 200 networks under different models with in-sample and out-sample realizations respectively. The number in parenthesis represents the standard deviation. First two columns represent the setup cost $\boldsymbol{f}$ and the holding cost $\boldsymbol{h}$. *In-sample* means that demand realizations in the simulation are drawn from the genuine gaussian distribuiton while *Out-sample* implies demand realizations are generated according to the mixture distribution described above. $\boldsymbol{p}$ is the penalty cost, or equivalently, the number of unsatisfied demand. *total cost* is the aggregated cost. It is clear that Model (CO) always not only achieves the lowest total cost but also has the smallest

Table 2: Cost Comparison, $(m, n) = (8, 8), (\rho, c_v, \kappa) = (0.3, 0.3, 1.0)$

			In-sample		Out-sample	
	f	**h**	**p**	total cost	**p**	total cost
SAA	67.7	660.2	64.5	792.5	70.0	798.0
	(14.9)	(70.5)	(31.7)	(78.2)	(38.8)	(81.0)
CO	63.4	668.5	53.1	785.0	57.3	789.2
	(13.9)	(66.5)	(12.2)	(78.5)	(15.0)	(82.0)
MV	63.3	636.1	96.1	795.5	108.7	808.1
	(13.7)	(66.1)	(15.4)	(80.1)	(15.8)	(81.3)

standard deviation. One remarkable decreasing appears in $\boldsymbol{p}$. It decreases from about 64.5 (by SAA) to 53.1 (by CO), i.e. 17.7% decreasing in unmet demand under the in-sample simulation. For out-sample simulation, the decreasing is also astounding, from 70.0 to 57.3, about 18.4%. Although the absolutely decreasing value is marginal, the relative decreasing is remarkable, which underlines the advatanges of our proposed conic model and reformulation technique: we can further significantly reduced the amount of unmet demand even it is already small. This managerial insight is quiet attractive, especially to disaster management experts, or fulfillment-oriented warehouse managers, since their first priority is reducing unmet demand.

We next scrutinize the unmet demand in a deeper way. Figure 2 shows the histograms about average unmet demand, where the horizontal axis shows the average unmet demand obtaining through simulations, and the vertical axis represents the number of networks falling in the corresponding bins. In the left graph,

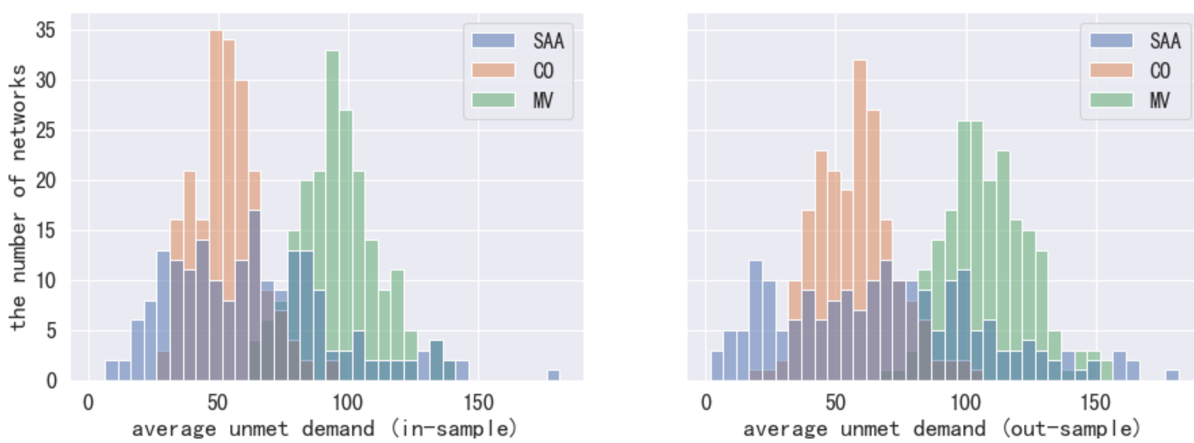

Figure 2: Histogram on average unmet demand, $(\rho, c_v, \kappa) = (0.3, 0.3, 1.0)$

when the true demand distribution aligns with the distribution that decision makers know, it is easy to see that Model (CO) almost dominates Model (MV) except some instances overlapping at aroud 75. Although Model (CO) achieves a higher level of concentration aroud 60, the least amount of unmet demand is achieved by Model (SAA) which is more scattered, entailing several worst cases at 150 levels. The right graph shows the results from out-sample simulations. Similarly, Model (CO) dominates Model (MV) most of the time. However, the most interesting change is the increasing dispersion of Model (SAA). Clear to see, the performance of Model (SAA) becomes unstable in different networks while (CO) and (MV) still are concentrated in their former mean values. The changes in standard deviations of $\boldsymbol{p}$ in Table 2 also verify our findings.

Furthur more, we calculate the 95%, 90%, 85% and 80% quantile of unmet demand and fulfillment rate to show the benefit of adopting Model (CO). When the true distribuiton is known to decision makers, these quantiles could be regarded as estimators of the value-at-risk of unmet demand after implementing the corresponding first-stage solutions. Table

Table 3: Unmet Demand with Various Quantiles

	In-sample				Out-sample			
quantile	95%	90%	85%	80%	95%	90%	85%	80%
SAA	406.4	212.8	111.0	56.3	335.4	325.9	250.0	119.8
CO	350.9	163.2	71.8	27.5	267.9	265.0	224.8	97.5
MV	617.8	365.5	206.2	100.2	569.7	569.7	398.3	171.5

4.2.2. Sensitivity Analysis

we modify some factors (correlation parameter ρ, coefficient of variation c_v, and risk attitude κ) to conduct sensitivity analysis within the same network.

4.3. General Network

4.4. Specific Network Structure

1. JG 1995 2. k-chain

total cost			In-sample Simulation									Out-sample Simulation						
			(6, 6)			(8, 8)			(10, 10)			(6, 6)			(8, 8)			
(rho, cv, kappa)			SAA	CO	MV	SAA	CO	MV	SAA	CO	MV	SAA	CO	MV	SAA	CO	MV	SA
0.30	0.05	1.0																
0.30	0.10	1.0																
0.30	0.15	1.0																
0.30	0.20	1.0																
0.30	0.25	1.0																
0.30	0.30	1.0																
0.30	0.35	1.0																
0.30	0.40	1.0																
0.30	0.45	1.0																
0.30	0.50	1.0																
0.05	0.30	1.0																
0.10	0.30	1.0																
0.15	0.30	1.0																
0.20	0.30	1.0																
0.25	0.30	1.0																
0.30	0.30	1.0																
0.35	0.30	1.0																
0.40	0.30	1.0																
0.45	0.30	1.0																
0.50	0.30	1.0																
0.30	0.30	0.5																
0.30	0.30	0.6																
0.30	0.30	0.7																
0.30	0.30	0.8																
0.30	0.30	0.9																
0.30	0.30	1.0																
0.30	0.30	1.1																
0.30	0.30	1.2																
0.30	0.30	1.3																
0.30	0.30	1.4																
0.30	0.30	1.5																

5. Extensions

5.1. k independent materials

5.2. Uncertainty in First- and Second-moments

6. Conclusion

References

Bertsimas, D., Delarue, A., Martin, S., 2019a. Optimizing schools' start time and bus routes. Proceedings of the National Academy of Sciences 116 (13), 5943–5948.

Bertsimas, D., Doan, X. V., Natarajan, K., Teo, C.-P., 2010. Models for minimax stochastic linear optimization problems with risk aversion. Mathematics of Operations Research 35 (3), 580–602.

Bertsimas, D., Sim, M., Zhang, M., 2019b. Adaptive distributionally robust optimization. Management Science 65 (2), 604–618.

Dickinson, P., 2010. An improved characterisation of the interior of the completely positive cone. The Electronic Journal of Linear Algebra 20.

Jordan, W. C., Graves, S. C., 1995. Principles on the benefits of manufacturing process flexibility. Management Science 41 (4), 577–594.

Kong, Q., Lee, C.-Y., Teo, C.-P., Zheng, Z., 2013. Scheduling arrivals to a stochastic service delivery system using copositive cones. Operations research 61 (3), 711–726.

Kong, Q., Li, S., Liu, N., Teo, C.-P., Yan, Z., 2020. Appointment scheduling under time-dependent patient no-show behavior. Management Science.

Mete, H. O., Zabinsky, Z. B., 2010. Stochastic optimization of medical supply location and distribution in disaster management. International Journal of Production Economics 126 (1), 76–84.

Natarajan, K., Teo, C. P., Zheng, Z., 2011. Mixed 0-1 linear programs under objective uncertainty: A completely positive representation. Operations research 59 (3), 713–728.

Ni, W., Shu, J., Song, M., 2018. Location and emergency inventory pre-positioning for disaster response operations: Min-max robust model and a case study of yushu earthquake. Production and Operations Management 27 (1), 160–183.

Yan, Z., Gao, S. Y., Teo, C. P., 2018. On the design of sparse but efficient structures in operations. Management Science 64 (7), 3421–3445.

7. Model Formulation and Reformulation

8. coding model

$$\exists M_1 \succeq_{sdp} 0, M_2 \geq \mathbf{0} \implies M = M_1 + M_2$$

Suppose

$$M = \begin{pmatrix} \tau & \boldsymbol{\xi}^T & \boldsymbol{\varphi}^T \\ \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{\psi}^T \\ \boldsymbol{\varphi} & \boldsymbol{\psi} & \boldsymbol{w} \end{pmatrix} = \begin{pmatrix} \tau & \boldsymbol{\xi}^T & \boldsymbol{\varphi}^T \\ \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{\psi}^T \\ \boldsymbol{\varphi} & \boldsymbol{\psi} & \boldsymbol{w} \end{pmatrix} + \begin{pmatrix} a & \boldsymbol{b}^T & \boldsymbol{c}^T \\ \boldsymbol{b} & \boldsymbol{e} & \boldsymbol{d}^T \\ \boldsymbol{c} & \boldsymbol{d} & \boldsymbol{f} \end{pmatrix} = M_1 + M_2$$

$\implies$

$$\begin{aligned}
L_{CO}(\boldsymbol{I}) = & \min_{\boldsymbol{I},\boldsymbol{Z},\alpha_i,\beta_i,\theta_j,\tau,\boldsymbol{\xi},\boldsymbol{\varphi},\boldsymbol{\psi},\boldsymbol{\eta},\boldsymbol{w}} \boldsymbol{f}^T\boldsymbol{I} + \boldsymbol{h}^T\boldsymbol{Z} + \sum_{i\in[M]} b_i\alpha_i + b_i^2\beta_i + (\tau + a) + 2\boldsymbol{\mu}^T(\boldsymbol{\xi} + \boldsymbol{b}) + \boldsymbol{\Sigma} \bullet (\boldsymbol{\eta} + \boldsymbol{e}) \\
& s.t. \sum_{i\in[M]} \boldsymbol{a}_i\alpha_i - \boldsymbol{\theta} - 2(\boldsymbol{\varphi} + \boldsymbol{c}) = \begin{bmatrix} \mathbf{0}_{n\times 1} \\ -\boldsymbol{I}_{m\times 1} \\ \mathbf{0}_{M\times 1} \end{bmatrix} \\
& \sum_{i\in[M]} \boldsymbol{a}_i\boldsymbol{a}_i^T\beta_i + \boldsymbol{Diag(\theta)} - (\boldsymbol{w} + \boldsymbol{f}) = \mathbf{0} \\
& -2(\boldsymbol{\psi} + \boldsymbol{d}) = \begin{bmatrix} \boldsymbol{e}^{[n]} \\ \mathbf{0} \end{bmatrix} \\
& \begin{pmatrix} \tau & \boldsymbol{\xi}^T & \boldsymbol{\varphi}^T \\ \boldsymbol{\xi} & \boldsymbol{\eta} & \boldsymbol{\psi}^T \\ \boldsymbol{\varphi} & \boldsymbol{\psi} & \boldsymbol{w} \end{pmatrix} \succeq_{sdp} \mathbf{0}, \begin{pmatrix} a & \boldsymbol{b}^T & \boldsymbol{c}^T \\ \boldsymbol{b} & \boldsymbol{e} & \boldsymbol{d}^T \\ \boldsymbol{c} & \boldsymbol{d} & \boldsymbol{f} \end{pmatrix} \geq 0
\end{aligned}$$

Appendix A. Proofs

Appendix A.1. Proof for Proposition 1

Proof. We prove this proposition by showing $g(\boldsymbol{I},\boldsymbol{Z},\tilde{\boldsymbol{d}}) \geq g(\boldsymbol{I},\tilde{\boldsymbol{d}})$ and $g(\boldsymbol{I},\boldsymbol{Z},\tilde{\boldsymbol{d}}) \leq g(\boldsymbol{I},\tilde{\boldsymbol{d}})$ both hold. For clearer notations, we rewrite both problems as follows:

$$\begin{aligned} g(\boldsymbol{I},\boldsymbol{Z},\tilde{\boldsymbol{d}}) = \quad \min_{\boldsymbol{x}} \quad & \sum_{j\in[n]} \tilde{d}_j - \sum_{(i,j)\in\mathcal{E}(\boldsymbol{Z})} x_{ij} \\ s.t. \quad & \sum_{i:(i,j)\in\mathcal{E}(\boldsymbol{Z})} x_{ij} \leq \tilde{d}_j, \forall j \in [n] \\ & \sum_{j:(i,j)\in\mathcal{E}(\boldsymbol{Z})} x_{ij} \leq I_i, \forall i \in [m] \\ & x_{ij} \geq 0, \forall (i,j) \in \mathcal{E}(\boldsymbol{Z}) \end{aligned} \tag{A.1}$$

Define a indices set $\boldsymbol{Z}^+$ containing all location indices that a warehouse is built, i.e. $\boldsymbol{Z}^+ := \{i \,\big|Z_i = 1, \ \forall i \in [m]\}$.

$$\begin{aligned} g(\boldsymbol{I},\tilde{\boldsymbol{d}}) = \quad \min_{\boldsymbol{x}} \quad & \sum_{j\in[n]} \tilde{d}_j - \sum_{(i,j)\in\mathcal{E}(\boldsymbol{Z})} x_{ij} - \sum_{(i,j)\notin\mathcal{E}(\boldsymbol{Z})} x_{ij} && \text{(A.2)} \\ s.t. \quad & \sum_{i:(i,j)\in\mathcal{E}(\boldsymbol{Z})} x_{ij} + \sum_{i:(i,j)\notin\mathcal{E}(\boldsymbol{Z})} x_{ij} \leq \tilde{d}_j, \forall j \in [n] && \text{(A.3)} \\ & \sum_{j:(i,j)\in\mathcal{E}} x_{ij} \leq I_i, \forall i \in \boldsymbol{Z}^+ \\ & \sum_{j:(i,j)\in\mathcal{E}} x_{ij} \leq 0, \forall i \notin \boldsymbol{Z}^+ && \text{(A.4)} \\ & x_{ij} \geq 0, \forall (i,j) \in \mathcal{E} \end{aligned}$$

In $g(\boldsymbol{I},\tilde{\boldsymbol{d}})$, the right-hand-side coefficient of the third constraint is zero since only establishing a warehouse is the prerequisite to store products there.

1. $g(\boldsymbol{I},\boldsymbol{Z},\tilde{\boldsymbol{d}}) \geq g(\boldsymbol{I},\tilde{\boldsymbol{d}})$

   It is easy to see that the optimal solution $\boldsymbol{x}^*$ to (A.1) is always a feasible solution to (A.2) by setting other $x_{ij} = 0, (i,j) \notin \mathcal{E}(\boldsymbol{Z})$. Therefore, $g(\boldsymbol{I},\boldsymbol{Z},\tilde{\boldsymbol{d}}) \geq g(\boldsymbol{I},\tilde{\boldsymbol{d}})$ holds.

2. $g(\boldsymbol{I},\boldsymbol{Z},\tilde{\boldsymbol{d}}) \leq g(\boldsymbol{I},\tilde{\boldsymbol{d}})$

   Now, suppose $\boldsymbol{x}^*$ is the optimal solution to (A.2). We split the solution into two parts according to whether $i \in \boldsymbol{Z}^+$, i.e. $\boldsymbol{x}^* = (\boldsymbol{x}^+, \boldsymbol{x}^0)$ where $\boldsymbol{x}^+$ contains allocation decision from warehouses while $\boldsymbol{x}^0$ contains allocation decisions from places no warehouse is built. Because of Constraint (A.4), we have $\boldsymbol{x}^0 = \boldsymbol{0}$. It is easy to check $\boldsymbol{x}^+$ is a feasible solution to (A.1). Therefore, $g(\boldsymbol{I},\boldsymbol{Z},\tilde{\boldsymbol{d}}) \leq g(\boldsymbol{I},\tilde{\boldsymbol{d}})$ always holds.

Combining above two statements, we finished the proof. □

Appendix A.2. Proof for Proposition 2

Proof. We prove Proposition 2 by equivalently reformulating (3) into a minimizing problem with only one constraint. And conducting optimization on the unique constraint has been proved to be NP-hard in Bertsimas et al. (2010).

Firstly, we take dual on (3) as follows:

$$
\begin{aligned}
L(\boldsymbol{I}) = \sup_{\mathbb{P}\in\mathscr{F}(\boldsymbol{\mu},\boldsymbol{\Sigma})} \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{I},\tilde{\boldsymbol{d}})] = \quad & \sup_{m(\tilde{\boldsymbol{d}})} \int_{\mathbb{R}^n_+} g(\boldsymbol{I},\tilde{\boldsymbol{d}})\ m(\tilde{\boldsymbol{d}}) \\
& s.t. \int_{\mathbb{R}^n_+} 1\ m(\tilde{\boldsymbol{d}}) = 1 \\
& \qquad \int_{\mathbb{R}^n_+} \tilde{\boldsymbol{d}}\ m(\tilde{\boldsymbol{d}}) = \boldsymbol{\mu} \\
& \qquad \int_{\mathbf{R}^n_+} \tilde{\boldsymbol{d}}\tilde{\boldsymbol{d}}^T\ m(\tilde{\boldsymbol{d}}) = \boldsymbol{\Sigma} \\
& \qquad m(\tilde{\boldsymbol{d}}) \in \mathscr{M}_+(\mathbb{R}^n_+)
\end{aligned}
$$

where $m(\tilde{\boldsymbol{d}})$ is naturally a probability measure on $\mathbb{R}^n_+$. The dual problem is:

$$
\begin{aligned}
& \inf_{s,\boldsymbol{t},\boldsymbol{\eta}} s + \boldsymbol{\mu}^T\boldsymbol{t} + \boldsymbol{\Sigma}\bullet\boldsymbol{\eta} \\
& s.t. \min_{(\hat{\boldsymbol{u}},\boldsymbol{v})\in\hat{\mathcal{L}_{lp}},\tilde{\boldsymbol{d}}\in\mathbb{R}^n_+} \tilde{\boldsymbol{d}}\boldsymbol{\eta}\tilde{\boldsymbol{d}} + \tilde{\boldsymbol{d}}^T(\boldsymbol{t}-\hat{\boldsymbol{u}}) + \boldsymbol{I}^T\boldsymbol{v} + s \geq 0
\end{aligned} \tag{A.5}
$$

Obviously, strong duality holds between the maximizing problem and the dual minimizing problem. According to Bertsimas et al. (2010) Theorem 3.1, the separation problem, i.e. for given $\boldsymbol{t}$, $\boldsymbol{I}$, and s, check if the constraint (A.5) is satisfied and if not, find a feasible $(\hat{\boldsymbol{u}},\boldsymbol{v})\in\hat{\mathcal{L}_{lp}}, \tilde{\boldsymbol{d}}\in\mathbb{R}^n_+$ satisfying (A.5), is NP-Complete. Because of the equivalence of separation and optimization, the minimizing problem is NP-hard, as well as evaluating the value of $L(\boldsymbol{I})$. □

Appendix A.3. Proof for Proposition 3

Appendix A.4. Proof for Proposition 4

Suppose the recourse decision $\boldsymbol{x} = \{x_{ij}\}_{i\in[m],j\in[n]}$ is an affine function based on the uncertain demand realization $\tilde{\boldsymbol{d}}$, e.g. $x_{ij} = \alpha_0^{ij} + \sum_{l\in[n]} \alpha_l^{ij}\tilde{d}_l$. For short, let us denote the recourse decision under affine function by $x_{ij}(\tilde{\boldsymbol{d}})$. Immediately, the second-stage worst-case expected value becomes.

$$
\begin{aligned}
L_{LDR}(\boldsymbol{I}) = \min \sup_{\mathbb{P} \in \mathscr{F}(\boldsymbol{\mu}, \boldsymbol{\Sigma})} & \sum_{j \in [n]} \tilde{d}_j - \sum_{(i,j) \in \mathcal{E}} x_{ij}(\tilde{\boldsymbol{d}}) \\
s.t. & \sum_{i \in \Gamma(j)} x_{ij}(\tilde{\boldsymbol{d}}) \leq \tilde{d}_j, \forall j \in [n], \forall \tilde{\boldsymbol{d}} \in \mathbb{R}_+^n \\
& \sum_{j \in \Gamma(i)} x_{ij}(\tilde{\boldsymbol{d}}) \leq I_i, \forall i \in [m], \forall \tilde{\boldsymbol{d}} \in \mathbb{R}_+^n \\
& \boldsymbol{x} \in \mathcal{LF}^r
\end{aligned}
$$

where

$$
\mathcal{LF}^r = \left\{ \boldsymbol{x} \in \mathbb{R}^{n,r} \;\middle|\; \begin{array}{l} \exists \alpha_0^{ij}, \alpha_l^{ij}, \forall i \in [m], \forall j \in [n], \forall l \in [n] : \\ x_{ij}(\tilde{\boldsymbol{d}}) = \alpha_0 + \sum_{l \in [n]} \alpha_l^{ij} \tilde{d}_l \end{array} \right\}
$$

Incorporating auxiliary variables as proposed by Bertsimas et al. (2019b), we can further modify the affine function as to provide more flexibility. To make the model tractable, we focus on a conservative approximation of the ambiguity set by only considering partial cross moment as shown in Bertsimas et al. (2019b):

$$
\begin{aligned}
L_{LDR}^{\partial}(\boldsymbol{I}) = \min \sup_{\mathbb{P} \in \mathscr{F}_{PCM}(\boldsymbol{\mu}, \boldsymbol{\Sigma})} & \sum_{j \in [n]} \tilde{d}_j - \sum_{(i,j) \in \mathcal{E}} x_{ij}(\tilde{\boldsymbol{d}}, \tilde{\boldsymbol{u}}) \\
s.t. & \sum_{i \in \Gamma(j)} x_{ij}(\tilde{\boldsymbol{d}}, \tilde{\boldsymbol{u}}) \leq \tilde{d}_j, \forall j \in [n], \forall (\tilde{\boldsymbol{d}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}} \\
& \sum_{j \in \Gamma(i)} x_{ij}(\tilde{\boldsymbol{d}}, \tilde{\boldsymbol{u}}) \leq I_i, \forall i \in [m], \forall (\tilde{\boldsymbol{d}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}} \\
& \boldsymbol{x} \in \mathcal{LF}^{r+K}
\end{aligned}
$$

where

$$
\mathcal{LF}^{r+K} = \left\{ \boldsymbol{x} \in \mathbb{R}^{n,r} \;\middle|\; \begin{array}{l} \exists \alpha_0^{ij}, \alpha_l^{ij}, \beta_k^{ij}, \forall i \in [m], \forall j \in [n], \forall l \in [n], \forall k \in [K] : \\ x_{ij}(\tilde{\boldsymbol{d}}, \tilde{\boldsymbol{u}}) = \alpha_0 + \sum_{l \in [n]} \alpha_l^{ij} \tilde{d}_l + \sum_{k \in [K]} \beta_k^{ij} \tilde{u}_k \end{array} \right\}
$$

and

$$
\mathscr{F}_{PCM}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \left\{ \mathbb{P} \in \mathcal{P}\left(\mathbb{R}_+^n\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{d}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{d}}] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}[\boldsymbol{f}_k^T \tilde{\boldsymbol{d}} \tilde{\boldsymbol{d}}^T \boldsymbol{f}_k] = \boldsymbol{f}_k^T \boldsymbol{\Sigma} \boldsymbol{f}_k, \ \forall k \in [K] \\ \mathbb{P}[\tilde{\boldsymbol{d}} \in \mathbb{R}_+^n] = 1 \end{array} \right\}
$$

Please note that, $\mathscr{F}_{PCM}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ asymptotically converge to $\mathscr{F}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as K goes to infinity.

Based on linear decision rule and partial cross moment ambiguity set, the prepositioning network design problem becomes:

$$
\begin{aligned}
Z_{LDR}^{\partial} = \min_{\boldsymbol{Z},\boldsymbol{I}} & \ \boldsymbol{f}^T\boldsymbol{Z} + \boldsymbol{h}^T\boldsymbol{I} + L_{LDR}^{\partial}(\boldsymbol{I}) \\
s.t. & \boldsymbol{I} \le M\boldsymbol{Z} \\
& \boldsymbol{I} \ge \boldsymbol{0}, \boldsymbol{Z} \in \{0,1\}^m
\end{aligned}
$$

Proposition 5. *Z_{LDR}^{∂} is equivalent to the following second order cone problem.*

$$
\begin{aligned}
& \min \boldsymbol{f}^T\boldsymbol{Z} + \boldsymbol{h}^T\boldsymbol{I} + \boldsymbol{\mu}^T\boldsymbol{t} + \sum_{k\in[K]} s_k \boldsymbol{f}_k^T \boldsymbol{\Sigma} \boldsymbol{f}_k + \gamma \\
s.t. & \gamma + \sum_{(i,j)\in\mathcal{E}} \alpha_0^{ij} \ge \frac{1}{2}\sum_{k\in[K]}(\varphi_k - \psi_k) \\
& \frac{1}{2}(\varphi_k + \psi_k) = \sum_{(i,j)\in\mathcal{E}} (\beta_k^{ij} + s_k), \ \forall k \in [K] \\
& \xi_k \boldsymbol{f}_k \le \boldsymbol{t} + \boldsymbol{\alpha}^{*,*} - \boldsymbol{1}, \ \forall k \in [K] \\
& -\sum_i \alpha_0^{ij} \ge \frac{1}{2}\sum_{k\in[K]}(\eta_k^j - \tau_k^j), \ \forall j \in [n] \\
& -\frac{1}{2}(\eta_k^j + \tau_k^j) = \sum_{i\in[m]} \beta_k^{ij}, \ \forall k \in [K], \forall j \in [n] \\
& -\theta_k^j \boldsymbol{f}_k \ge \boldsymbol{\alpha}^{*,j}, \ \forall k \in [K], \forall j \in [n] \\
& I_i - \sum_{j\in[n]} \alpha_0^{ij} \ge \frac{1}{2}\sum_{k\in[K]}(w_k^i - \lambda_k^i), \ \forall i \in [m] \\
& -\frac{1}{2}(w_k^i + \lambda_k^i) = \sum_{j\in[n]} \beta_k^{ij}, \ \forall k \in [K], \forall i \in [m] \\
& -\delta_k^i \boldsymbol{f}_k \ge \boldsymbol{\alpha}^{i,*}, \ \forall k \in [K], \forall i \in [m] \\
& \boldsymbol{I} \le M\boldsymbol{Z}, \boldsymbol{I} \ge \boldsymbol{0}, \boldsymbol{Z} \in \{0,1\}^n, s_k \ge 0, \forall k \in [K] \\
& \begin{bmatrix} \varphi_k \\ \psi_k \\ \xi_k \end{bmatrix} \succeq_{SOC} \boldsymbol{0} \ \forall k \in [K], \begin{bmatrix} \eta_k^j \\ \tau_k^j \\ \theta_k^j \end{bmatrix} \succeq_{SOC} \boldsymbol{0} \ \forall k \in [K] \ \forall j \in [n], \begin{bmatrix} w_k^i \\ \lambda_k^i \\ \delta_k^i \end{bmatrix} \succeq_{SOC} \boldsymbol{0} \ \forall k \in [K] \ \forall i \in [m]
\end{aligned}
$$

where $\boldsymbol{I}, \boldsymbol{Z}, \boldsymbol{t}, \gamma, s_k, \varphi_k, \psi_k, \xi_k, \eta_k^j, \tau_k^j, \theta_k^j, w_k^i, \lambda_k^i, \delta_k^i, \alpha_0^{ij}, \alpha_l^{ij}, \beta_k^{ij}$ are decision variables. In addition, $\boldsymbol{\alpha}^{*,*} =$

$$\begin{bmatrix} \sum_{(i,j)\in\mathcal{E}} \alpha_1^{ij} \\ \vdots \\ \sum_{(i,j)\in\mathcal{E}} \alpha_n^{ij} \end{bmatrix}, \boldsymbol{\alpha}^{*,j} = \begin{bmatrix} \sum_{i\in[m]} \alpha_1^{ij} \\ \vdots \\ \sum_{i\in[m]} \alpha_j^{ij} - 1 \\ \vdots \\ \sum_{i\in[m]} \alpha_n^{ij} \end{bmatrix}, \text{and } \boldsymbol{\alpha}^{i,*} = \begin{bmatrix} \sum_{j\in[n]} \alpha_1^{ij} \\ \vdots \\ \sum_{j\in[n]} \alpha_n^{ij} \end{bmatrix}$$