Journal of Industrial and Management Optimization

JIMO, 22(10): 4796–4824.
DOI: 10.3934/jimo.2026166
Received: 27 March 2026
Revised: 03 September 2026
Accepted: 04 September 2026
Published: 09 September 2026

https://www.aimspress.com/journal/jimo

Research article

A cross-regional emergency supply pre-positioning with partial collaboration cost-sharing strategy

Jingke Zhou[1] and Yingzhen Chen[2,*]

[1] School of Mathematics and Statistics, Hubei University of Arts and Science, Xiangyang, Hubei, 441053, China
[2] School of Economics and Management, Hubei University of Arts and Science, Xiangyang, Hubei, 441053, China

* **Correspondence:** Email: yzchen@hbuas.edu.cn.

Appendix

Appendix A: Detailed proof of Theorem 1

This appendix provides a step-by-step derivation of the mixed-integer linear reformulation given in Theorem 1. The proof follows the framework established in Xie [1] for distributionally robust chance constraints with right-hand uncertainty.

1. From the distributionally robust chance constraint to a worst-expectation form

The original constraint requires $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\tilde{d}_{it} \leq x_i^t, \forall i \in N, t \in T\right) \geq 1-\epsilon$. For any probability distribution $\mathbb{P}$, $\mathbb{P}\left(\tilde{d}_{it} \leq x_i^t, \forall i, t\right) = 1 - \mathbb{P}\left(\exists i, t\colon \tilde{d}_{it} > x_i^t\right)$. Therefore, the constraint is equivalent to

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\exists i, t\colon \tilde{d}_{it} > x_i^t\right) \leq \epsilon \tag{A.1}$$

Define the indicator of violation $\mathbb{I}(\boldsymbol{x},\tilde{\boldsymbol{d}}):=\mathbb{I}(\exists i,t\colon \tilde{d}_{it}>x_i^t)=\max_{i\in N,t\in T}\mathbb{I}(\tilde{d}_{it}>x_i^t)$. Because the probability of an event equals the expectation of its indicator, we have $\mathbb{P}(\exists i,t\colon \tilde{d}_{it}>x_i^t)=\mathbb{E}_{\mathbb{P}}[\mathbb{I}(\boldsymbol{x},\tilde{\boldsymbol{d}})]$. Thus, (A.1) becomes

$$\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left[\max_{i,t}\mathbb{I}(\tilde{d}_{it}>x_i^t)\right]\le\epsilon \tag{A.2}$$

This formulation is now in the standard form required for applying Wasserstein duality.

2. Application of Wasserstein strong duality

Let $\ell(\boldsymbol{x},\boldsymbol{d}):=\mathbb{I}(\boldsymbol{x},\boldsymbol{d})$. The function $\ell(\boldsymbol{x},\cdot)$ is bounded (it takes only the values 0 and 1) and upper semicontinuous (the set where it equals 1 is open). For such a loss, the strong duality theorem for the Wasserstein ambiguity set [2,3] states that

$$\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}[\ell(x,\tilde{\boldsymbol{d}})]=\min_{\lambda\ge0}\left\{\lambda\delta+\frac{1}{|K|}\sum_{k\in K}\sup_{d\in\Xi}\left[\ell(x,d)-\lambda\|\boldsymbol{d}-\hat{\boldsymbol{d}}^k\|\right]\right\} \tag{A.3}$$

Outline of the duality proof.

The result follows from the Kantorovich dual representation of the Wasserstein distance and a standard Lagrangian duality/minimax argument. Introducing a Lagrange multiplier $\lambda\ge0$ for the constraint $W(\mathbb{P},\mathbb{P}_{\hat{d}})\le\delta$ and using $W(\mathbb{P},\mathbb{P}_{\hat{\boldsymbol{d}}})=\sup_{\|h\|_{Lip}\le1}\left\{\int h d\mathbb{P}-\frac{1}{|K|}\sum_{k\in K}h(\hat{\boldsymbol{d}}^k)\right\}$, one exchanges the order of supremum and infimum via Sion's minimax theorem (the loss is bounded and the space of 1-Lipschitz functions is compact in an appropriate topology). The inner supremum over $\mathbb{P}$ reduces to a pointwise supremum because $\mathbb{P}$ can be a Dirac measure. Finally, the optimization over Lipschitz functions is solved by the Moreau-Yosida regularization, yielding the explicit term $sup_d[\ell(x,d)-\lambda\|\boldsymbol{d}-\hat{\boldsymbol{d}}^k\|]$. A detailed proof can be found in the references above; here, we only need the final formula (A.3).

Applying (A.3) to $\ell=\mathbb{I}$ and converting the supremum to a negative infimum gives

$$\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}[\mathbb{I}(x,\tilde{\boldsymbol{d}})]=\min_{\lambda\ge0}\left\{\lambda\delta-\frac{1}{|K|}\sum_{k\in K}\inf_{d}\left[\lambda\|\boldsymbol{d}-\hat{\boldsymbol{d}}^k\|-\mathbb{I}(x,d)\right]\right\} \tag{A.4}$$

Combining (A.2) with (A.4), we obtain the necessary and sufficient condition

$$\exists\lambda\ge0\colon\lambda\delta-\frac{1}{|K|}\sum_{k\in K}\inf_{\boldsymbol{d}}\left[\lambda\|\boldsymbol{d}-\hat{\boldsymbol{d}}^k\|-\mathbb{I}(\boldsymbol{x},\boldsymbol{d})\right]\le\epsilon \tag{A.5}$$

3. Explicit evaluation of the infimum term

Fix a sample index $k\in K$ and set $\mathcal{J}_k(x,\lambda):=\inf_{d}\left[\lambda\|\boldsymbol{d}-\hat{\boldsymbol{d}}^k\|-\mathbb{I}(x,d)\right]$. Because $\mathbb{I}(x,d)=\max_{i,t}\mathbb{I}(d_{it}>x_i^t)$, we use the following elementary identity (Xie [1], Lemma 1): $\inf_{d}\Big[f_0(d)+$

$\min_{i,t} f_{it}(d)\Big] = \min_{i,t} \inf_{d} [f_0(d) + f_{it}(d)]$, valid for any functions f_0, f_{it} (the infima are taken in the extended real sense). Taking $f_0(d) = \lambda \|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\|$ and $f_{it}(d) = -\mathbb{I}(d_{it} > x_i^t)$ yields

$$\mathcal{J}_k(x, \lambda) = \min_{i \in N, t \in T} \inf_{d} \left[\lambda \|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\| - \mathbb{I}(d_{it} > x_i^t)\right] \tag{A.6}$$

Now, analyze the inner infimum for a fixed pair (i, t). Split the domain according to the value of the indicator.

Region 1: $d_{it} > x_i^t$.

Here, $\mathbb{I}(d_{it} > x_i^t) = 1$, so the objective is $\lambda \|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\| - 1$. The minimal distance to this region is obtained by increasing the coordinate d_{it} just above x_i^t while keeping all other coordinates at their sample values. Hence, $\inf_{d: d_{it} > x_i^t} \|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\| = \max\{x_i^t - \hat{d}_{it}^k, 0\}$.

Region 2: $d_{it} \le x_i^t$.

Here, $\mathbb{I}(d_{it} > x_i^t) = 0$, so the objective is $\lambda \|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\|$. If $\hat{d}_{it}^k \le x_i^t$, the point $\widehat{\boldsymbol{d}}^k$ itself is feasible, giving distance 0; otherwise, the closest point is obtained by decreasing d_{it} to x_i^t . Thus, $\inf_{d: d_{it} \le x_i^t} \|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\| = \max\{\hat{d}_{it}^k - x_i^t, 0\}$.

Consequently,

$$\inf_{\boldsymbol{d}} \left[\lambda \|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\| - \mathbb{I}(d_{it} > x_i^t)\right] = \min\left\{\lambda \max\{x_i^t - \hat{d}_{it}^k, 0\} - 1, \lambda \max\{\hat{d}_{it}^k - x_i^t, 0\}\right\} \tag{A.7}$$

We simplify (A.7) by a case analysis.

Case 1: $\hat{d}_{it}^k \le x_i^t$.

Then, $\max\{x_i^t - \hat{d}_{it}^k, 0\} = x_i^t - \hat{d}_{it}^k \ge 0$ and $\max\{\hat{d}_{it}^k - x_i^t, 0\} = 0$. The right-hand side of (A.7) becomes $\min\{\lambda(x_i^t - \hat{d}_{it}^k) - 1, 0\}$.

Case 2: $\hat{d}_{it}^k > x_i^t$.

Then, $\max\{x_i^t - \hat{d}_{it}^k, 0\} = 0$ and $\max\{\hat{d}_{it}^k - x_i^t, 0\} = \hat{d}_{it}^k - x_i^t > 0$. Now, the right-hand side is $\min\{-1, \lambda(\hat{d}_{it}^k - x_i^t)\} = -1$, because the second term is non-negative. Notice that in Case 2, the value is -1, which equals $\min\{\lambda \cdot 0 - 1, 0\}$, and in Case 1, it is $\min\{\lambda(x_i^t - \hat{d}_{it}^k) - 1, 0\}$. Therefore, in both cases, we can write uniformly

$$\inf_{\boldsymbol{d}} \left[\lambda \|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\| - \mathbb{I}(d_{it} > x_i^t)\right] = \min\{\lambda \max\{x_i^t - \hat{d}_{it}^k, 0\} - 1, 0\} \tag{A.8}$$

Insert (A.8) into (A.6). Using the property $\min_{i,t} \min\{a_{it}, 0\} = \min\left\{\min_{i,t} a_{it}, 0\right\}$, we obtain

$$\mathcal{J}_k(x, \lambda) = \min\left\{\lambda \min_{i \in N, t \in T} \max\{x_i^t - \hat{d}_{it}^k, 0\} - 1, 0\right\} \tag{A.9}$$

Define $\Phi_k(\boldsymbol{x}) := \min_{i \in N, t \in T} \max\{x_i^t - \hat{d}_{it}^k, 0\}$. Then (A.9) becomes

$$\mathcal{J}_k(x,\lambda) = min\{\lambda \Phi_k(\boldsymbol{x}) - 1, 0\} \tag{A.10}$$

4. Variable transformation and simplification

Substituting (A.10) into (A.5) gives

$$\lambda\delta - \frac{1}{|K|}\sum_{k\in K} min\,\{\lambda\Phi_k(\boldsymbol{x}) - 1{,}0\} \le \epsilon \tag{A.11}$$

Using the identity $min\{a,0\} = -max\{-a,0\}$, we rewrite (A.11) as

$$\lambda\delta + \frac{1}{|K|}\sum_{k\in K} max\,\{1 - \lambda\Phi_k(\boldsymbol{x}), 0\} \le \epsilon \tag{A.12}$$

If $\lambda = 0$, then (A.12) reduces to $1 \le \epsilon$, which contradicts $\epsilon \in (0{,}1)$. Hence, $\lambda > 0$, and we can set $\gamma := 1/\lambda > 0$. Multiplying (A.12) by γ and using the homogeneity of the max operator,

$$\delta - \epsilon\gamma + \frac{1}{|K|}\sum_{k\in K} max\,\{\gamma - \Phi_k(\boldsymbol{x}), 0\} \le 0 \tag{A.13}$$

Now define for each $k \in K$, $v_k := -max\{\gamma - \Phi_k(\boldsymbol{x}), 0\} = min\{\Phi_k(\boldsymbol{x}) - \gamma, 0\} \le 0$.

Then, (A.13) simplifies to the linear inequality

$$\delta - \epsilon\gamma \le \frac{1}{|K|}\sum_{k\in K} v_k \tag{A.14}$$

5. Linearization with auxiliary variables

We now construct a mixed-integer linear system that captures the definition of v_k. Introduce for each $k \in K$ a continuous variable $s_k \ge 0$ and a binary variable $u_k \in \{0{,}1\}$.

The intended meaning is $s_k = max\{\Phi_k(\boldsymbol{x}), 0\}$, $u_k = \begin{cases} 1 & if\ \Phi_k(\boldsymbol{x}) > 0, \\ 0 & otherwise. \end{cases}$

Because $\Phi_k(\boldsymbol{x})$ is already non-negative (it is a minimum of non-negative terms), we actually have $s_k = \Phi_k(\boldsymbol{x})$. The auxiliary variables merely help to linearize.

The relation $v_k = min\{s_k - \gamma, 0\}$ is enforced by

$$v_k + \gamma \le s_k,\ v_k \le 0 \tag{A.15}$$

Why this works:

If $s_k - \gamma \ge 0$, the tightest upper bound on v_k from (A.15) is 0 (since $v_k \le 0$ is stronger than $v_k \le s_k - \gamma$). If $s_k - \gamma < 0$, the bound $v_k \le s_k - \gamma$ is the active one. In an optimization problem, the variables v_k appear on the right-hand side of (A.14); larger v_k make the constraint easier to satisfy, so at optimality, v_k will attain its maximum feasible value, which is exactly $\min\{s_k - \gamma, 0\}$.

To enforce $s_k = \Phi_k(\boldsymbol{x})$, we use the big-M linearization of the minimum of affine functions (Xie [1], Corollary 3). Observe that $\Phi_k(\boldsymbol{x}) = \underset{i,t}{min} max\{x_i^t - \hat{d}_{it}^k, 0\}$.

The following system, together with the binary variable u_k, is equivalent:

$$s_k \le x_i^t - \hat{d}_{it}^k + M(1 - u_k), \forall i \in N, t \in T \tag{A.16}$$

$$s_k \le M u_k \tag{A.17}$$

$$s_k \ge 0 \tag{A.18}$$

where M is a sufficiently large constant satisfying $M \ge \underset{i\in N, t\in T, k\in K}{max} \left| x_i^t - \hat{d}_{it}^k \right|$ for all feasible $\boldsymbol{x}$.

Verification of correctness.

If $u_k = 0$: (A.17) forces $s_k = 0$. Then, (A.16) becomes $0 \le x_i^t - \hat{d}_{it}^k + M$, which is automatically true because the right-hand side is at least $M - max|\cdot| \ge 0$. In this case, we must have $\Phi_k(\boldsymbol{x}) = 0$. Otherwise, if $\Phi_k(\boldsymbol{x}) > 0$, the solver could set $u_k = 1$ to allow a positive s_k. Hence, at optimality, $u_k = 0$ implies $\Phi_k(\boldsymbol{x}) = 0 = s_k$.

If $u_k = 1$: (A.17) imposes no effective upper bound. Constraints (A.16) give $s_k \le x_i^t - \hat{d}_{it}^k + M$; these are always true. The only lower bound on s_k comes from (A.18). However, s_k also appears in (A.15). Because larger s_k allow larger v_k (which is beneficial for feasibility), the optimal solution will not increase s_k beyond the necessary value. The smallest value compatible with u_k=1 is exactly $\Phi_k(\boldsymbol{x})$. Consequently, at optimality, we have $s_k = \Phi_k(\boldsymbol{x})$.

However, this reasoning relies on the fact that the optimization drives v_k to its maximum feasible value. To obtain an equivalent formulation that holds regardless of the optimization direction, we supplement (A.15) with the following lower bounds.

The original constraints $v_k + \gamma \le s_k$, $v_k \le 0$ only provide an upper bound. To enforce equality without relying on the optimization direction, we introduce an additional binary variable $b_k \in \{0,1\}$, and use the same big-M constant. The correct linearization is:

$$v_k \ge s_k - \gamma - M(1 - b_k) \tag{A.19}$$

$$v_k \ge -Mb_k \tag{A.20}$$

$$b_k \in \{0,1\} \tag{A.21}$$

Together with (A.15), we obtain a complete set of constraints that exactly force $v_k = min\{s_k - \gamma, 0\}$.

Verification.

Case $s_k - \gamma \ge 0$. The true value is $v_k = 0$. Set $b_k = 0$. Then, (A.19) gives $v_k \ge s_k - \gamma - M$, which is automatically satisfied. (A.20) gives $v_k \ge 0$. Together with $v_k \le 0$ from (A15), we obtain $v_k = 0$.

Case $s_k - \gamma < 0$. The true value is $v_k = s_k - \gamma$. Set $b_k = 1$. Then, (A.20) becomes $v_k \ge -M$, which is not binding. (A.19) gives $v_k \ge s_k - \gamma$. Together with the upper bounds $v_k \le s_k - \gamma$ and $v_k \le 0$ from (A.15), we obtain $v_k = s_k - \gamma$.

Thus, the system (A.15)–(A.21) exactly enforces $v_k = min\{s_k - \gamma, 0\}$ for any feasible solution.

6. *Final mixed-integer linear system*

Collecting (A.14), (A.15), and (A.16)–(A.21), we obtain the complete reformulation stated in Theorem 1:

$$Z = \left\{ x \in \mathbb{R}^n : \begin{array}{c} \delta - \epsilon\gamma \le \dfrac{1}{|K|}\displaystyle\sum_{k \in K} v_k, \\ v_k + \gamma \le s_k, \forall k \in K, \\ s_k \le x_i^t - \hat{d}_{it}^k + M(1 - u_k), \forall k \in K, i \in N, t \in T, \\ s_k \le Mu_k, \forall k \in K, \\ v_k \ge s_k - \gamma - M(1 - b_k), \forall k \in K, \\ v_k \ge -Mb_k, \forall k \in K, \\ s_k \ge 0, \gamma \ge 0, v_k \le 0, u_k \in \{0,1\}, b_k \in \{0,1\}, \forall k \in K. \end{array} \right\}$$

This completes the detailed proof. □

The preceding steps invoked the Wasserstein strong duality theorem, exchanging infimum and minimum, the equivalence of the linearization, and the appropriate choice of the big-M constant. The following four subsections provide complete proofs of these technical results, thereby filling in the details that were only cited or briefly sketched in the main derivation.

A.1 Proof of Wasserstein strong duality (Lemma 1)

Lemma 1. Under the assumptions that (i) the loss function $\ell(\boldsymbol{x},\cdot)$ is upper semicontinuous and bounded above, and (ii) the Wasserstein ambiguity set $\mathcal{P}$ is defined as in (3), the following equality holds:

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\ell\left(x,\tilde{d}\right)\right] = \min_{\lambda\geq 0}\left\{\lambda\delta + \frac{1}{|K|}\sum_{k\in K}\sup_{d\in\Xi}\left[\ell(x,d) - \lambda\left\|\boldsymbol{d}-\widehat{\boldsymbol{d}}^k\right\|\right]\right\} \tag{A.22}$$

Proof. We provide a complete derivation using Lagrangian duality, Kantorovich duality, and minimax theory.

The worst-expectation problem is $\mathcal{J}(x) := \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\ell\left(x,\tilde{d}\right)\right] = \sup_{\mathbb{P}}\left\{\int_{\Xi}\ell(x,d)\mathbb{P}(dd) : W(\mathbb{P},\mathbb{P}_{\hat{d}}) \leq \delta\right\}$.

Introduce a Lagrange multiplier $\lambda \geq 0$ for the constraint $W(\mathbb{P},\mathbb{P}_{\hat{d}}) \leq \delta$. The Lagrangian is $L(\mathbb{P},\lambda) = \int_{\Xi}\ell\,(x,d)\mathbb{P}(d\boldsymbol{d}) + \lambda\left(\delta - W(\mathbb{P},\mathbb{P}_{\hat{d}})\right)$.

By weak duality, $\mathcal{J}(x) \leq \inf_{\lambda\geq 0}\sup_{\mathbb{P}} L(\mathbb{P},\lambda)$.

For any two distributions $\mathbb{P},\mathbb{Q}$, $W(\mathbb{P},\mathbb{Q}) = \sup_{\|h\|_{Lip}\leq 1}\left\{\int_{\Xi}h(d)\mathbb{P}(d\boldsymbol{d}) - \int_{\Xi}h(d)\mathbb{Q}(d\boldsymbol{d})\right\}$, where $\|h\|_{Lip} = \sup_{d_1\neq d_2}|h(d_1)-h(d_2)|/\|\boldsymbol{d_1}-\boldsymbol{d_2}\|$. Applying this to $W(\mathbb{P},\mathbb{P}_{\hat{d}})$ with $\mathbb{P}_{\hat{d}} = \frac{1}{|K|}\sum_{k\in K}\delta_{\hat{d}^k}$ yields

$$W(\mathbb{P},\mathbb{P}_{\hat{d}}) = \sup_{\|h\|_{Lip}\leq 1}\left\{\int_{\Xi} h(d)\mathbb{P}(dd) - \frac{1}{|K|}\sum_{k\in K}h\left(\hat{d}^k\right)\right\} \tag{A.23}$$

Insert (A.23) into the Lagrangian: $L(\mathbb{P},\lambda) = \int_{\Xi}\ell(\boldsymbol{x},\boldsymbol{d})\mathbb{P}(d\boldsymbol{d}) + \lambda\delta - \lambda\sup_{\|h\|_{Lip}\leq 1}\left\{\int_{\Xi}\ell(\boldsymbol{x},\boldsymbol{d})\mathbb{P}(d\boldsymbol{d}) - \frac{1}{|K|}\sum_{k\in K}h\left(\widehat{\boldsymbol{d}}^k\right)\right\}$.

Taking supremum over $\mathbb{P}$ and infimum over λ,

$$\mathcal{J}(\boldsymbol{x}) \leq \inf_{\lambda\geq 0}\left\{\lambda\delta + \sup_{\mathbb{P}}\left[\int_{\Xi}\ell(\boldsymbol{x},\boldsymbol{d})\mathbb{P}(d\boldsymbol{d}) - \lambda\sup_{\|h\|_{Lip}\leq 1}\left\{\int_{\Xi}h(\boldsymbol{d})\mathbb{P}(d\boldsymbol{d}) - \frac{1}{|K|}\sum_{k\in K}h\left(\widehat{\boldsymbol{d}}^k\right)\right\}\right]\right\} \tag{A.24}$$

For a fixed $\mathbb{P}$, the term involving $\sup_{\|h\|_{Lip}\leq 1}$ is convex in h. Because $\ell(\boldsymbol{x},\cdot)$ is bounded above and upper semicontinuous, and the set of 1-Lipschitz functions on a Polish space is compact in the topology of uniform convergence on compact sets, the conditions of Sion's minimax theorem are satisfied [4]. Hence, we may exchange $\sup_{\mathbb{P}}$ and $\sup_{\|h\|_{Lip}\leq 1}$:

$$\sup_{\mathbb{P}} \inf_{\|h\|_{Lip}\le 1} \left\{\int_{\Xi} [\ell(\boldsymbol{x},\boldsymbol{d}) - \lambda h(\boldsymbol{d})]\, \mathbb{P}(d\boldsymbol{d}) + \frac{\lambda}{|K|}\sum_{k\in K} h\left(\widehat{\boldsymbol{d}}^k\right)\right\} = \inf_{\|h\|_{Lip}\le 1} \sup_{\mathbb{P}} \left\{\int_{\Xi} [\ell(\boldsymbol{x},\boldsymbol{d}) - \lambda h(\boldsymbol{d})]\mathbb{P}(d\boldsymbol{d}) + \frac{\lambda}{|K|}\sum_{k\in K} h\left(\widehat{\boldsymbol{d}}^k\right)\right\} \quad (A.25)$$

Computing the inner supremum over $\mathbb{P}$. For a fixed 1-Lipschitz function h, $\sup_{\mathbb{P}} \int_{\Xi} [\ell(\boldsymbol{x},\boldsymbol{d}) - \lambda h(\boldsymbol{d})]\, \mathbb{P}(d\boldsymbol{d}) = \sup_{\boldsymbol{d}\in\Xi} [\ell(\boldsymbol{x},\boldsymbol{d}) - \lambda h(\boldsymbol{d})]$, because a Dirac measure concentrated at the point where the function attains its supremum (or any point arbitrarily close to it) achieves the value. Thus, (A.22) becomes

$$\inf_{\|h\|_{Lip}\le 1} \left\{\sup_{\boldsymbol{d}\in\Xi}[\ell(\boldsymbol{x},\boldsymbol{d}) - \lambda h(\boldsymbol{d})] + \frac{\lambda}{|K|}\sum_{k\in K} h\left(\widehat{\boldsymbol{d}}^k\right)\right\} \quad (A.26)$$

This is a functional optimization problem. Define for each sample $k \in K$, $\phi_k(\boldsymbol{d}) := \ell(\boldsymbol{x},\boldsymbol{d}) - \lambda\|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\|$.

A standard result in optimal transport states that for any 1-Lipschitz h, $\ell(\boldsymbol{x},\boldsymbol{d}) - \lambda h(\boldsymbol{d}) \le \sup_{\boldsymbol{d}'\in\Xi}\left[\ell(\boldsymbol{x},\boldsymbol{d}') - \lambda\left\|\boldsymbol{d}' - \widehat{\boldsymbol{d}}^k\right\|\right] + \lambda h\left(\widehat{\boldsymbol{d}}^k\right)$, $\forall \boldsymbol{d} \in \Xi$ [5,6]. Taking the supremum over $\boldsymbol{d}$ and summing over k, $\sup_{\boldsymbol{d}\in\Xi}[\ell(\boldsymbol{x},\boldsymbol{d}) - \lambda h(\boldsymbol{d})] + \frac{\lambda}{|K|}\sum_{k\in K} h\left(\widehat{\boldsymbol{d}}^k\right) \ge \frac{1}{|K|}\sum_{k\in K} \sup_{\boldsymbol{d}\in\Xi}\left[\ell(\boldsymbol{x},\boldsymbol{d}) - \lambda\left\|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\right\|\right]$.

Moreover, this lower bound is attained by the specific 1-Lipschitz function $h^*(\boldsymbol{d}) = \min_{k\in K}\left\{\frac{1}{\lambda}\left(\sup_{d'\in\Xi}\left[\ell(\boldsymbol{x},\boldsymbol{d}') - \lambda\left\|\boldsymbol{d}' - \widehat{\boldsymbol{d}}^k\right\|\right] + \lambda\left\|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\right\|\right)\right\}$, which can be verified to be 1-Lipschitz and yields equality. Therefore, the infimum in (A.26) equals

$$\frac{1}{|K|}\sum_{k\in K} \sup_{\boldsymbol{d}\in\Xi}\left[\ell(\boldsymbol{x},\boldsymbol{d}) - \lambda\left\|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\right\|\right] \quad (A.27)$$

Establishing strong duality. Substituting (A.24) back, we obtain the upper bound $\mathcal{J}(\boldsymbol{x}) \le \inf_{\lambda\ge 0}\left\{\lambda\delta + \frac{1}{|K|}\sum_{k\in K} \sup_{\boldsymbol{d}\in\Xi}\left[\ell(\boldsymbol{x},\boldsymbol{d}) - \lambda\left\|\boldsymbol{d} - \widehat{\boldsymbol{d}}^k\right\|\right]\right\}$.

To prove the reverse inequality, take any feasible $\mathbb{P} \in \mathcal{P}$ and a coupling $\mathbb{Q}$ between $\mathbb{P}$ and $\mathbb{P}_{\hat{d}}$ with $\int\|\boldsymbol{d} - \boldsymbol{d}'\|d\mathbb{Q} \le \delta$. For any $\lambda \ge 0$, $\mathbb{E}_{\mathbb{P}}\left[\ell\left(\boldsymbol{x},\widetilde{\boldsymbol{d}}\right)\right] = \int_{\Xi\times\Xi} \ell\,(\boldsymbol{x},\boldsymbol{d})\mathbb{Q}(d\boldsymbol{d},d\boldsymbol{d}') \le \int_{\Xi\times\Xi} \left[\sup_{\boldsymbol{d}''}[\ell(\boldsymbol{x},\boldsymbol{d}'') - \lambda\|\boldsymbol{d}'' - \boldsymbol{d}'\| + \lambda\|\boldsymbol{d} - \boldsymbol{d}'\|]\right]\mathbb{Q}(d\boldsymbol{d},d\boldsymbol{d}') = \frac{1}{|K|}\sum_{k\in K} \sup_{\boldsymbol{d}''}\left[\ell(\boldsymbol{x},\boldsymbol{d}'') - \lambda\left\|\boldsymbol{d}'' - \widehat{\boldsymbol{d}}^k\right\| + \lambda\delta\right.$

Taking supremum over $\mathbb{P}$ on the left and infimum over λ on the right gives $\mathcal{J}(\boldsymbol{x}) \ge$, the right-hand side of (A.22). Hence, equality holds. □

A.2 Proof of the inf-min exchange lemma (Lemma 3.1).

Lemma 3.1. For any functions $f_0: \Xi \to \mathbb{R} \cup \{+\infty\}$ and $\left\{f_{i,t}: \Xi \to \mathbb{R} \cup \{+\infty\}\right\}_{i\in N, t\in T}$, $\inf_{\boldsymbol{d}\in\Xi}\left[f_0(\boldsymbol{d}) + \min_{i,t} f_{it}(\boldsymbol{d})\right] = \min_{i,t}\inf_{\boldsymbol{d}\in\Xi}[f_0(\boldsymbol{d}) + f_{it}(\boldsymbol{d})]$, where the infima are taken in the extended real sense.

Proof. Let $m := \underset{i,t}{min}\, \underset{\boldsymbol{d}}{inf}[f_0(\boldsymbol{d}) + f_{it}(\boldsymbol{d})]$. For any $\boldsymbol{d}$, $f_0(\boldsymbol{d}) + \underset{i,t}{min} f_{it}(\boldsymbol{d}) \leq f_0(\boldsymbol{d}) + f_{it}(\boldsymbol{d})\ \ \forall i,t.$ Taking infimum over $\boldsymbol{d}$ on the left and then minimum over i, t on the right yields

$$\underset{\boldsymbol{d}}{inf}\left[f_0(\boldsymbol{d}) + \underset{i,t}{min} f_{it}(\boldsymbol{d})\right] \leq m \tag{A.28}$$

For the reverse inequality, fix $\epsilon > 0$. By definition of the infimum, there exists $\boldsymbol{d}_\epsilon$ such that $f_0(\boldsymbol{d}_\epsilon) + \underset{i,t}{min} f_{it}(\boldsymbol{d}_\epsilon) \leq \underset{\boldsymbol{d}}{inf}\left[f_0(\boldsymbol{d}) + \underset{i,t}{min} f_{it}(\boldsymbol{d})\right] + \epsilon.$

Let (i_ϵ, t_ϵ) be indices achieving the minimum at $\boldsymbol{d}_\epsilon$ (if the minimum is not attained, choose a sequence approaching it; the argument extends by a limiting procedure). Then, $f_0(\boldsymbol{d}_\epsilon) + f_{i_\epsilon t_\epsilon}(\boldsymbol{d}_\epsilon) = f_0(\boldsymbol{d}_\epsilon) + \underset{i,t}{min} f_{it}(\boldsymbol{d}_\epsilon) \leq \underset{d}{inf}\left[f_0(\boldsymbol{d}) + \underset{i,t}{min} f_{it}(\boldsymbol{d})\right] + \epsilon.$

Now take infimum over $\boldsymbol{d}$ on the left-hand side for the specific pair (i_ϵ, t_ϵ) : $\underset{\boldsymbol{d}}{inf}\left[f_0(\boldsymbol{d}) + f_{i_\epsilon t_\epsilon}(\boldsymbol{d})\right] \leq f_0(\boldsymbol{d}_\epsilon) + f_{i_\epsilon t_\epsilon}(\boldsymbol{d}_\epsilon) \leq \underset{\boldsymbol{d}}{inf}\left[f_0(\boldsymbol{d}) + \underset{i,t}{min} f_{it}(\boldsymbol{d})\right] + \epsilon.$

Since $\epsilon > 0$ was arbitrary, we obtain

$$\min_{i,t} \inf_{\boldsymbol{d}} [f_0(\boldsymbol{d}) + f_{it}(\boldsymbol{d})] \leq \inf_{\boldsymbol{d}}\left[f_0(\boldsymbol{d}) + \min_{i,t} f_{it}(\boldsymbol{d})\right] \tag{A.29}$$

Combining (A.28) and (A.29) gives equality. □

A.3 Proof of linearization equivalence.

Claim. The mixed-integer linear system (A.15)–(A.21), together with (A.14), correctly enforces $s_k = \Phi_k(\boldsymbol{x})$ and $v_k = \min\{\Phi_k(\boldsymbol{x}) - \gamma, 0\}$ at optimality.
Proof. We prove the equivalence by constructing an optimal solution with the desired properties.
Let $(\boldsymbol{x}^*, \gamma^*, \boldsymbol{v}^*, \boldsymbol{s}^*, \boldsymbol{u}^*)$ be any optimal solution of the overall optimization problem [minimizing or maximizing an objective subject to (A.14)–(A.21) and the main inequality (A.13)]. Define $\Phi_k^* := \Phi_k(\boldsymbol{x}^*) = \underset{i,t}{min}\, max\{x_i^{t*} - \hat{d}_{it}^k, 0\}$.

Consistency of the binary variables u_k. If $\Phi_k^* > 0$, then any feasible solution must have $u_k^* = 1$; otherwise, with $u_k^* = 0$, (A.17) forces $s_k^* = 0$ and (A.18) is satisfied, but then (A.14) would give $v_k^* \leq -\gamma^*$. However, because $\Phi_k^* > 0$, the true value $v_k = min\{\Phi_k^* - \gamma^*, 0\}$ is strictly larger than $-\gamma^*$ when $\gamma^* < \Phi_k^*$, and the objective (or feasibility) would be improved by taking $u_k = 1$ and allowing $s_k > 0$. Hence, in an optimal solution, we must have $u_k^* = 1$ whenever $\Phi_k^* > 0$. Conversely, if $\Phi_k^* = 0$, setting $u_k^* = 0$ is feasible and yields; any solution with $u_k^* = 1$ would allow $s_k^* > 0$, but increasing s_k^* cannot improve feasibility [it only relaxes (A.14)], so there exists an optimal solution with $u_k^* = 0$ and $s_k^* = 0$. Without loss of generality, we may assume that the optimal solution satisfies $u_k^* = 1$ iff $\Phi_k^* > 0$.

Construction of a solution with $s_k^* = \Phi_k^*$. Suppose in the optimal solution we have $s_k^* > \Phi_k^*$ for some k. Consider a modified solution where we replace s_k^* by Φ_k^* while keeping all other variables unchanged. We check feasibility:

(A.14): $v_k^* + \gamma^* \le s_k^*$ becomes tighter when s_k^* is reduced; since the original inequality held and $\Phi_k^* < s_k^*$, the new inequality is tighter but may be violated if v_k^* was already at its upper bound. However, from the definition $v_k = min\{\Phi_k(\boldsymbol{x}) - \gamma, 0\}$, the correct value of v_k^* should satisfy $v_k^* \le \Phi_k^* - \gamma^*$. If the original v_k^* violated this, it could be decreased (made more negative) without affecting the objective if the objective depends favorably on larger v_k [as they appear on the right-hand side of (A.14)]. In fact, because v_k appears positively in (A.14), the optimal solution will have v_k^* as large as possible subject to $v_k^* \le s_k^* - \gamma^*$ and $v_k^* \le 0$. Reducing s_k^* to Φ_k^* may force a reduction in v_k^* to maintain $v_k^* \le \Phi_k^* - \gamma^*$. But any such reduction can be compensated by increasing other v_j (if possible) or by adjusting; if no adjustment is possible, the original solution could not have been optimal because the constraint (A.14) would be tighter than necessary. A rigorous argument: define $\tilde{v}_k = min\{\Phi_k^* - \gamma^*, 0\} \le v_k^*$. Replacing v_k^* by $\tilde{v}_k$ and keeping all other variables unchanged, we obtain a feasible solution with a possibly smaller left-hand side in (A.14). If the original solution was optimal, we must have $v_k^* = min\{\Phi_k^* - \gamma^*, 0\}$. Hence, $v_k^* \le \Phi_k^* - \gamma^*$ and consequently, $v_k^* + \gamma^* \le \Phi_k^*$. Therefore, reducing s_k^* to Φ_k^* does not violate (A.14). All other constraints are unaffected. Thus, the modified solution is feasible.

Optimality of the modified solution. The modification either keeps v_k^* unchanged (if $\tilde{v}_k = v_k^*$) or reduces it. Reducing v_k^* makes (A.14) harder to satisfy, which cannot improve the objective if the objective is monotone in the direction of feasibility. In a maximization problem where larger v_k are beneficial, reducing v_k would worsen the objective; hence, such a move would not be optimal unless forced. Therefore, in an optimal solution, we must have $s_k^* = \Phi_k^*$ and $v_k^* = min\{\Phi_k^* - \gamma^*, 0\}$.

The constraints (A.19)–(A.21), together with (A.15), provide both upper and lower bounds on v_k. As shown in the main proof, for any feasible (s_k, γ), the system forces $v_k = min\{s_k - \gamma, 0\}$ exactly, without relying on the optimization direction. This completes the proof that the linearized system is equivalent to the original condition. □

A.4 Choice and validity of the big-M constant.

Proposition. Let the decision variables be bounded, e.g., $x_i^t \in [L, U]$ for all i, t. Define $M := \max_{i\in N, t\in T, k\in K} max\{|U - \hat{d}_{it}^k|, |L - \hat{d}_{it}^k|\}$. Then, for every feasible $\boldsymbol{x}$, $|x_i^t - \hat{d}_{it}^k| \le M \ \forall i, t, k$.

Moreover, with this choice of M, the big-M constraints (A.16)–(A.17) are valid (i.e., they do not cut off any feasible solution) and the linearization remains exact.

Proof. The bound follows directly from $x_i^t \in [L, U]$.

Validity for u_k constraints.

When $u_k = 0$: (A.17) forces $s_k = 0$. The constraint $s_k \le x_i^t - \hat{d}_{it}^k + M$ becomes $0 \le x_i^t - \hat{d}_{it}^k + M$. Since $x_i^t - \hat{d}_{it}^k \ge -M$ by construction, the right-hand side is non-negative, so the constraint is automatically satisfied.

When $u_k = 1$: (A.16) becomes $s_k \le x_i^t - \hat{d}_{it}^k + M$. Because s_k will be at most $\Phi_k(\boldsymbol{x}) \le max_{i,t}(x_i^t - \hat{d}_{it}^k) \le M$, the right-hand side is at least $-M + M = 0$, and the inequality holds trivially.

Validity for b_k constraints.

Recall the additional constraints introduced for the rigorous linearization of $v_k = min\{s_k - \gamma, 0\}$.

If $b_k = 0$: (A.20) gives $v_k \ge 0$. Together with $v_k \le 0$ from (A.15), we obtain $v_k = 0$. The term $-M(1 - b_k) = -M$ in (A.19) is not binding because the lower bound 0 is higher. This case corresponds to $s_k - \gamma \geq 0$, and the chosen M does not restrict feasibility since the active bound is $v_k \geq 0$, which is tight.

If $b_k = 1$: (A.20) gives $v_k \geq -M$, which is very weak. (A.19) gives $v_k \geq s_k - \gamma$. In practice, with reasonable ϵ and δ, the optimal γ remains moderate, and the chosen M works without additional tuning. Alternatively, one may replace the big-M terms in (A.19)–(A.20) by indicator constraints (available in commercial solvers like Gurobi), which eliminates the need for an explicit M and guarantees exactness without numerical issues. In our implementation, we use indicator constraints for the b_k variables.

Thus, with the chosen M (or a suitably large constant), the big-M constraints are valid and do not cut off any feasible solution. Using the smallest possible M improves the computational performance of the mixed-integer linear programming relaxation.

In practice, to balance numerical accuracy and efficiency, we determine a tight value of M as follows.

1). Solve a relaxed version of the original model where the distributionally robust chance constraint (7) is replaced by its expectation (i.e., using mean demands), and all binary variables (u_k, b_k) are relaxed to continuous in [0,1]. This linear program yields an upper bound $x_i^{t,UB}$ for each inventory variable x_i^t.

2). Compute $M = max_{i,t,k}\{|x_i^t - \hat{d}_{it}^k|\}$.

In our numerical instance, this procedure gave $M \approx 1.5 \times 10^5$. With this value, the big-M constraints (A.16) and (A.19)–(A.20) are tight enough to avoid numerical instability while still being valid (no feasible solution is cut off). We observed that using a larger M (e.g., 10^9) caused mild numerical issues for very small δ (e.g., $\delta = 0.001$), whereas the tight M obtained from the LP relaxation worked reliably for all tested δ.

If an even tighter bound is desired, one can iterate between solving the MILP and updating M, but in our experiments, a single LP pre-solve was sufficient.

□

References for appendix

1. W. Xie, On distributionally robust chance constrained programs with Wasserstein distance, *Math. Program.*, **186** (2021), 115–155. https://doi.org/10.1007/s10107-019-01445-5

2. J. Blanchet, K. Murthy, Quantifying distributional model risk via optimal transport, *Math. Oper. Res.*, **44** (2019), 565–600. https://doi.org/10.1287/moor.2018.0936

3. R. Gao, A. Kleywegt, Distributionally robust stochastic optimization with Wasserstein distance,*Math. Oper. Res.*, **48** (2023), 603–655. https://doi.org/10.1287/moor.2022.1275

4. M. Sion, On general minimax theorems, *Pac. J. Math.*, **8** (1958), 171–176. https://doi.org/10.2140/pjm.1958.8.171

5. C. Villani, *Optimal transport: Old and new*, Grundlehren Math. Wiss. 338, Berlin: Springer, 2009. https://doi.org/10.1007/978-3-540-71050-9

6. F. Santambrogio, *Optimal transport for applied mathematicians: Calculus of variations*, PDEs, and Modeling, Cham: Birkhäuser/Springer, 2015. https://doi.org/10.1007/978-3-319-20828-2

Appendix B: Detailed data

Table B1. Costs for each focal region under different defection scenarios.

	Partial collaboration	Defector					
		AH	HA	HB	HN	JX	SX
AH	1.36E+05	1.36E+05	1.22E+05	1.27E+05	1.20E+05	1.23E+05	1.23E+05
HA	1.36E+05	1.24E+05	4.18E+04	1.27E+05	1.01E+05	8.19E+04	8.22E+04
HB	1.36E+05	1.24E+05	9.55E+04	2.14E+05	1.20E+05	1.22E+05	1.23E+05
HN	1.36E+05	1.24E+05	1.22E+05	6.86E+04	1.57E+05	1.23E+05	1.23E+05
JX	1.36E+05	1.24E+05	1.22E+05	1.27E+05	1.20E+05	1.31E+05	1.23E+05
SX	1.58E+04	5.27E+04	9.22E+04	1.26E+05	1.20E+05	1.23E+05	2.53E+04
Total	6.97E+05	6.85E+05	5.94E+05	7.91E+05	7.39E+05	7.05E+05	6.00E+05

Table B2. Impact of disaster risk weight on risk-based costs of regions.

	AH	HA	HB	HN	JX	SX
0.1	7.3729E+04	2.0876E+05	1.4020E+05	3.4576E+04	9.1517E+04	2.3069E+05
0.2	7.3729E+04	1.8920E+05	1.3908E+05	4.0253E+04	8.9815E+04	2.0767E+05
0.3	7.3729E+04	1.6943E+05	1.3796E+05	4.5990E+04	8.8095E+04	1.8441E+05
0.4	7.3729E+04	1.4946E+05	1.3682E+05	5.1789E+04	8.6356E+04	1.6089E+05
0.5	7.3729E+04	1.2926E+05	1.3566E+05	5.7651E+04	8.4598E+04	1.3712E+05
0.6	7.3729E+04	1.0885E+05	1.3450E+05	6.3577E+04	8.2821E+04	1.1309E+05
0.7	8.4017E+04	1.0052E+05	1.5192E+05	7.9275E+04	9.2331E+04	1.0119E+05
0.8	1.1005E+05	1.0052E+05	1.9722E+05	1.1288E+05	1.1823E+05	9.5879E+04
0.9	1.6025E+05	1.0052E+05	2.8456E+05	1.7768E+05	1.6817E+05	8.5644E+04
1	2.9744E+05	1.0052E+05	5.2327E+05	3.5476E+05	3.0465E+05	5.7673E+04

Table B3. Optimal cost shares z_i under different Wasserstein radii δ.

	AH	HA	HB	HN	JX	SX
0.001	1.3198E+05	1.2976E+05	1.3615E+05	1.3495E+05	1.3095E+05	1.2217E+04
0.002	1.3199E+05	1.2976E+05	1.3616E+05	1.3496E+05	1.3096E+05	1.2218E+04
0.003	1.3199E+05	1.2977E+05	1.3616E+05	1.3496E+05	1.3097E+05	1.2218E+04
0.004	1.3157E+05	1.3022E+05	1.3617E+05	1.3497E+05	1.3096E+05	1.2219E+04
0.005	1.3158E+05	1.3022E+05	1.3618E+05	1.3498E+05	1.3096E+05	1.2219E+04
0.006	1.3158E+05	1.3023E+05	1.3618E+05	1.3498E+05	1.3097E+05	1.2220E+04
0.007	1.3159E+05	1.3024E+05	1.3619E+05	1.3499E+05	1.3098E+05	1.2221E+04
0.008	1.2921E+05	1.3222E+05	1.3620E+05	1.3500E+05	1.3083E+05	1.2784E+04
0.009	1.2921E+05	1.3223E+05	1.3620E+05	1.3500E+05	1.3084E+05	1.2783E+04
0.01	1.2922E+05	1.3223E+05	1.3621E+05	1.3501E+05	1.3084E+05	1.2782E+04
0.011	1.2923E+05	1.3224E+05	1.3622E+05	1.3501E+05	1.3085E+05	1.2782E+04
0.012	1.2924E+05	1.3225E+05	1.3622E+05	1.3502E+05	1.3086E+05	1.2781E+04
0.013	1.2924E+05	1.3225E+05	1.3623E+05	1.3503E+05	1.3086E+05	1.2780E+04
0.014	1.2925E+05	1.3219E+05	1.3624E+05	1.3503E+05	1.3087E+05	1.2852E+04
0.015	1.2940E+05	1.3126E+05	1.3624E+05	1.3504E+05	1.3095E+05	1.3581E+04
0.016	1.2940E+05	1.3126E+05	1.3625E+05	1.3505E+05	1.3095E+05	1.3582E+04
0.017	1.2941E+05	1.3127E+05	1.3625E+05	1.3505E+05	1.3096E+05	1.3583E+04
0.018	1.2941E+05	1.3128E+05	1.3626E+05	1.3506E+05	1.3096E+05	1.3584E+04
0.019	1.2942E+05	1.3128E+05	1.3627E+05	1.3507E+05	1.3097E+05	1.3585E+04
0.02	1.3092E+05	1.3010E+05	1.3627E+05	1.3507E+05	1.2809E+05	1.6166E+04

Table B4. Extended sensitivity analysis of δ (0.0001 to 0.1).

δ	0.0001	0.0003	0.001	0.003	0.01	0.03	0.1
Objective value	1.3614E+05	1.3615E+05	1.3615E+05	1.3616E+05	1.3621E+05	1.3634E+05	1.3680E+05