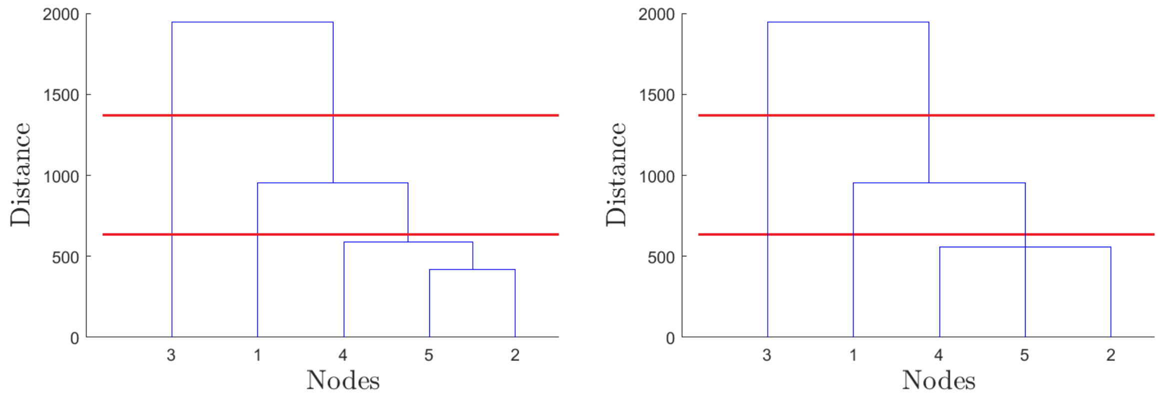
This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.



(a) (b)

Figure EC.1 Creating a nested fulfillment cost structure with $L = 4$ from a dendrogram.

EC.1. Algorithms

We collate the algorithms in our paper in this section.

EC.1.1. Approximating distance-based fulfillment costs with a nested structure

To read about the context of this algorithm, please refer to Section 2.1 of the paper.

Algorithm 1: Hierarchical Agglomerative Clustering Algorithm

Input: Distance matrix $\mathcal{R} = (r_{ij})_{i,j\in[n]}$, fulfillment cost function $\phi : \Re^+ \mapsto \Re^+$, and cluster distance function $L : 2^{[n]} \times 2^{[n]} \mapsto \Re$

Output: Nested fulfillment costs $\mathbf{s} = (s_{\ell,k})$ and structure $\Xi = \{\mathbf{E}_0, \mathbf{E}_1, \ldots, \mathbf{E}_{L-1}\}$ where $L = n$

```
S ← {{1}, {2}, ..., {n}}                                  // Clusters in current level
R̄ = (r̄_ij) ← R
E_0 ← I_n                                   // Set level 0 structure to identity matrix
for k ← 1 to n do
    s_{0,k} ← φ(r_kk)                                              // Set level 0 costs
for ℓ ← 1 to n − 1 do
    Choose the two closest nodes i*, j* = argmin_{i,j∈S} r̄_ij.
    Cluster i*, j* into a single node: S ← S + {i*, j*} − {i*} − {j*}
    // Recalculate the distance matrix
    forall I, J ∈ S do
        r̄_{I,J} ← L(I, J)
    for k ← 1 to |S| do
        I ← S(k)                                                  // k^th cluster of S
        Set k^th row of E_ℓ to be 0-1 where entry i is 0 iff i ∈ I
        s_{ℓ,k} ← φ(r̄_{II})
```

EC.1.2. Example for Generating Nested Fulfillment Cost Structures with $L < n$

Consider Figure EC.1, where a nested fulfillment cost structure with $L = 4$ $(n = 5)$ is created from the dendrogram in Figure 2a. Here, the range of distances are partitioned into three quantiles by the two lines drawn on the dendrogram. In Figure EC.1a, the lower line gives rise to three connected components: $\{\{3\}, \{1\}, \{4, 5, 2\}\}$. The nodes in each connected component are considered to be a single cluster in level $l = 1$, and the UPGMA distances are recalculated for the new clusters. The upper line gives rise to two connected components: $\{\{3\}, \{1, 4, 5, 2\}\}$, which form the two components at level $l = 2$, resulting in Figure EC.1b.

EC.1.3. Solving the robust multi-location newsvendor problem heuristically

The dual of the SDP (22) is a minimization problem. Therefore, from Lemma 1, an upper bound to the minmax cost C^* defined in (DRIP), is

$$
\begin{aligned}
\hat{C} := \min_{\substack{\mathbf{y},t_0,\mathbf{t},\mathbf{Y},\mathbf{u},\\ \mathbf{B},\mathbf{W},\mathbf{U},\mathbf{V}}} \quad & h\cdot\mathbf{e}^\top(\mathbf{y}-\mathbf{m}) + \mathbf{s}_0^\top\mathbf{m} + t_0 + \mathbf{t}^\top\mathbf{m} + \langle \mathbf{Y}, \boldsymbol{\Sigma} + \mathbf{m}\mathbf{m}^\top\rangle + \mathbf{e}^\top\mathbf{B}\mathbf{e} \\
\text{s.t.} \quad & \begin{pmatrix} t_0 & \frac{1}{2}\mathbf{t}^\top & \frac{1}{2}\mathbf{u}^\top \\ \frac{1}{2}\mathbf{t} & \mathbf{Y} & -\frac{1}{2}\mathbf{V}^\top \\ \frac{1}{2}\mathbf{u} & -\frac{1}{2}\mathbf{V} & \mathbf{U} \end{pmatrix} \succeq 0, \\
& \mathbf{u} = -\mathbf{W}\mathbf{e} + \left(\mathbf{B}+\mathbf{B}^\top\right)\mathbf{e} + \mathbf{P}\mathbf{y}, \\
& \mathbf{V} \geq \mathbf{P}, \\
& \mathbf{U} \leq \mathbf{W} - \mathbf{B}, \\
& \mathbf{W}, \mathbf{B} \geq 0, \\
& t_0 \in \Re,\ \mathbf{t}\in\Re^n,\ \mathbf{u}\in\Re^N,\ \mathbf{Y}\in\Re^{n\times n},\ \mathbf{B},\mathbf{W},\mathbf{U}\in\Re^{N\times N},\ \mathbf{V}\in\Re^{N\times n}.
\end{aligned}
\tag{EC.1}
$$

where $\mathbf{P} := \left(\mathbf{E}_{L-1}^\top \mathrm{diag}(\eta_{L-1})\ \mathbf{E}_{L-2}^\top \mathrm{diag}(\eta_{L-2})\ \cdots\ \mathbf{E}_0^\top \mathrm{diag}(\eta_0)\right)^\top \in \Re^{N\times n}$. Thus, a computationally tractable heuristic for finding robust inventory levels under a nested cost structure is to choose the inventory levels based on the optimal $\mathbf{y}$ from solving the SDP (EC.1). The procedure is outlined below.

Algorithm 2: Algorithm for Robust Heuristic Inventory Solution

Input: Demand vector mean $\mathbf{m}$ and covariance $\boldsymbol{\Sigma}$, penalty cost p, overage cost h, nested fulfillment costs $\mathbf{s} = (s_{\ell,k})$ and structure $\Xi = \{\mathbf{E}_0, \mathbf{E}_1, \ldots, \mathbf{E}_{L-1}\}$

Output: Heuristic solution $\hat{\mathbf{y}}$ to the robust multi-location newsvendor problem and approximate minmax robust cost $\hat{C}$

// Construct η_ℓ for $\ell \in [0, L-1]$

$\eta_{L-1} \leftarrow p + h - s_{L-1}$

for $\ell \leftarrow 1$ **to** $L-2$ **do**

  for $k \leftarrow 1$ **to** number of rows in $\mathbf{E}_\ell$ **do**

    $\mathcal{I} \leftarrow$ cluster corresponding to column indices where entry in $\mathbf{E}_{\ell,k}$ is equal to 1

    $m \leftarrow$ level $\ell+1$ index of parent node to $\mathcal{I}$

    $\eta_{\ell,k} \leftarrow s_{\ell+1,m} - s_{\ell,k}$

$\mathbf{P} \leftarrow \left(\mathbf{E}_{L-1}^\top \mathrm{diag}(\eta_{L-1})\ \mathbf{E}_{L-2}^\top \mathrm{diag}(\eta_{L-2})\ \cdots\ \mathbf{E}_0^\top \mathrm{diag}(\eta_0)\right)^\top$

Solve the SDP (EC.1) with data $\mathbf{m}, \boldsymbol{\Sigma}, \mathbf{P}, h, \mathbf{s}_0$

Set $\hat{\mathbf{y}}$ to be the optimal $\mathbf{y}$ from solving the SDP

Set $\hat{C}$ to be the optimal value of the SDP

EC.2. Proofs

Here, we collect the proofs of the paper.

EC.2.1. Proof of Lemma 1

To obtain the fulfillment cost of a demand realization $\mathbf{D}$ in an L-level nested fulfillment cost structure, we sum the fulfillment costs in each level. The total fulfillment cost in level 0 is $\sum_{i\in[n]} s_{0,i}\cdot\min(d_i, y_i) = \mathbf{s}_0^\top\mathbf{D} - \sum_{i\in[n]} s_{0,i}\cdot(d_i - y_i)^+$, where $\mathbf{s}_0 = (s_{0,i})_{i\in[n]}$. For $\ell = 1, 2, \ldots, L-1$, the total units of demand in locations of $\mathcal{I}_k^{(\ell)}$ fulfilled at level ℓ (at a per-unit fulfillment cost $s_{\ell,k}$) is

$$\underbrace{\sum_{m\in\mathcal{K}_k^{(\ell)}}\left(\sum_{i\in\mathcal{I}_m^{(\ell-1)}} d_i - \sum_{i\in\mathcal{I}_m^{(\ell-1)}} y_i\right)^+}_{\text{unmet demand in } \mathcal{I}_k^{(\ell)} \text{ after level } \ell-1} - \underbrace{\left(\sum_{i\in\mathcal{I}_k^{(\ell)}} d_i - \sum_{i\in\mathcal{I}_k^{(\ell)}} y_i\right)^+}_{\text{unmet demand in } \mathcal{I}_k^{(\ell)} \text{ after level } \ell},$$

where $\mathcal{K}_k^{(\ell)}$ are all level $\ell-1$ children of set $\mathcal{I}_k^{(\ell)}$. Since $p+h$ is strictly greater than all fulfillment costs, then the penalty cost is $p\cdot(\mathbf{e}^\top\mathbf{D}-\mathbf{e}^\top\mathbf{y})^+$, and the overage cost is $h\cdot(\mathbf{e}^\top\mathbf{y}-\mathbf{e}^\top\mathbf{D})^+$. Therefore, the total cost (overage, penalty and fulfillment) is equal to

$$\begin{aligned} C(\mathbf{y},\mathbf{D}) &= h\cdot\mathbf{e}^\top(\mathbf{y}-\mathbf{D}) + \mathbf{s}_0^\top\mathbf{D} + (p+h-s_{L-1})\cdot\left(\mathbf{e}^\top\mathbf{D}-\mathbf{e}^\top\mathbf{y}\right)^+ \\ &\quad + \sum_{\ell=0}^{L-2}\sum_{k\in[n_\ell]}(s_{\ell+1,m^{(\ell+1)}(k)} - s_{\ell,k})\cdot\left(\sum_{i\in\mathcal{I}_k^{(\ell)}} d_i - \sum_{i\in\mathcal{I}_k^{(\ell)}} y_i\right)^+, \end{aligned} \tag{EC.2}$$

where $m^{(\ell+1)}(k)\in[n_{\ell+1}]$ is the level $\ell+1$ parent of set $k\in[n_\ell]$. Using compact notation with the parameters η_ℓ and assignment matrices $\mathbf{E}_\ell$, we obtain the lemma. ■

EC.2.2. Proof of Lemma 2

We define

$$\begin{aligned} M(\mathbf{y}) = \max_f \quad & \mathbb{E}_f\left[\sum_{\ell=0}^{L-1}\eta_\ell^\top(\mathbf{E}_\ell\mathbf{D}-\mathbf{E}_\ell\mathbf{y})^+\right] \\ \text{s.t.} \quad & \mathbb{E}_f(1)=1, \\ & \mathbb{E}_f\left(\tilde{\mathbf{D}}\right)=\mathbf{m}, \\ & \mathbb{E}_f\left(\tilde{\mathbf{D}}\tilde{\mathbf{D}}^\top\right)=\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^\top, \\ & f(\mathbf{D})\ge 0, \quad \forall\mathbf{D}\in\Re^n. \end{aligned} \tag{EC.3}$$

which is equal to the left-hand side of (6).

Since $\Sigma\succ 0$, then the moments $(\mathbf{m},\mathbf{\Sigma})$ are strictly in the interior of the feasible moment cone. Hence, strong duality of moment problems holds (Smith 1995). The dual of the moment problem is

$$\begin{aligned} M(\mathbf{y}) = \min_{t,\mathbf{r},\mathbf{Y}} \quad & t+\mathbf{r}^\top\mathbf{m}+\langle\mathbf{Y},\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^\top\rangle \\ \text{s.t.} \quad & t+\mathbf{r}^\top\mathbf{x}+\mathbf{x}^\top\mathbf{Y}\mathbf{x}\ge\sum_{\ell=0}^{L-1}\eta_\ell^\top(\mathbf{E}_\ell\mathbf{x}-\mathbf{E}_\ell\mathbf{y})^+, \quad \forall\mathbf{x}\in\Re^n \end{aligned} \tag{EC.4}$$

We can reformulate the dual as the following semi infinite linear program:

$$\begin{aligned} M(\mathbf{y}) = \min_{t,\mathbf{r},\mathbf{Y}} \quad & t+\mathbf{r}^\top\mathbf{m}+\langle\mathbf{Y},\mathbf{\Sigma}+\mathbf{m}\mathbf{m}^\top\rangle \\ \text{s.t.} \quad & t+\mathbf{r}^\top\mathbf{x}+\mathbf{x}^\top\mathbf{Y}\mathbf{x}\ge\sum_{\ell=0}^{L-1}(\eta_\ell\odot\mathbf{e}_{A_\ell})^\top(\mathbf{E}_\ell\mathbf{x}-\mathbf{E}_\ell\mathbf{y}), \qquad \forall\mathbf{x}\in\Re^n, \\ & \qquad \forall(A_0,A_1,\cdots,A_{L-1})\in 2^{[n_0]}\times 2^{[n_1]}\times 2^{[n_{L-1}]} \end{aligned} \tag{EC.5}$$

where $\odot$ is the element-wise product operator, and $\mathbf{e}_{A_\ell}$ is an n_ℓ-dimensional binary vector whose k^{th} element is 1 if and only if $k\in A_\ell$. For simplicity, we can write the right-hand side as $\mathbf{a}_k^\top\mathbf{x}+\mathbf{b}_k^\top\mathbf{y}$ for $k\in[2^N]$, where $N=\sum_{\ell=0}^{L-1}n_\ell$. Note that $\mathbf{b}_k=-\mathbf{a}_k$. The constraint now becomes: $\mathbf{x}^\top\mathbf{Y}\mathbf{x}+(\mathbf{r}-\mathbf{a_k})^\top\mathbf{x}+t-\mathbf{b}_k^\top\mathbf{y}\ge 0,\forall\mathbf{x}$. This is true if and only if

$$\begin{bmatrix} \mathbf{Y} & \frac{1}{2}(\mathbf{r}-\mathbf{a_k}) \\ \frac{1}{2}(\mathbf{r}-\mathbf{a_k})^\top & t-\mathbf{b}_k^\top\mathbf{y} \end{bmatrix}\succeq 0, \quad \forall k. \quad ■$$

EC.2.3. Proof of Proposition 1

Let $M(y_1,y_2)$ be the optimal value of the moment problem

$$M(y_1,y_2) := \inf_{f\in\mathcal{F}} \mathbb{E}_f\left[\eta\min(d_1,y_1)+\eta\min(d_2,y_2)+\zeta\min(d_1+d_2,y_1+y_2)\right]. \tag{EC.6}$$

Using the relation $(a-b)^+ = a-\min(a,b)$, we have that

$$\sup_{f\in\mathcal{F}}\mathbb{E}_f\left[\eta(d_1-y_1)^+ + \eta(d_2-y_2)^+ + \zeta(d_1+d_2-y_1-y_2)^+\right] = (\eta+\zeta)(m_1+m_2) - M(y_1,y_2).$$

Therefore, to prove the proposition, it is sufficient to show

$$M(y_1,y_2) = \frac{1}{2}(\eta+\zeta)(y_1+y_2+m_1+m_2) - (\eta+\zeta)\Phi\left(\mathbf{y};\frac{3\zeta+\eta}{\zeta+\eta}\right)$$

where $\Phi(\cdot;\nu)$ is defined in (8).

Note that the dual of the moment problem (EC.6) we want to solve in the case where there are two non-identical regions is

$$\begin{aligned}
\sup_{t,u_1,u_2,r_1,r_2,v}\quad & t+m_1r_1+m_2r_2+(m_1^2+\sigma_1^2)u_1+(m_2^2+\sigma_2^2)u_2+(m_1m_2+\rho\sigma_1\sigma_2)v\\
\text{s.t.}\quad & t+r_1d_1+r_2d_2+u_1d_1^2+u_2d_2^2+vd_1d_2\\
&\quad \le \zeta\min(d_1+d_2,y_1+y_2)+\eta\min(d_1,y_1)+\eta\min(d_2,y_2), \qquad \forall (d_1,d_2)\in\Re^2.
\end{aligned} \tag{EC.7}$$

Define $g(d_1,d_2;t,u_1,u_2,r_1,r_2,v) := t+r_1d_1+r_2d_2+u_1d_1^2+u_2d_2^2+vd_1d_2$. Then we have

$$\begin{aligned}
\sup_{t,u,r,v}\quad & t+m_1r_1+m_2r_2+(m_1^2+\sigma_1^2)u_1+(m_2^2+\sigma_2^2)u_2+(m_1m_2+\rho\sigma_1\sigma_2)v\\
\text{s.t.}\quad & g(d_1,d_2;t,u_1,u_2,r_1,r_2,v)\le(\eta+\zeta)(d_1+d_2), && \forall d_1\le y_1,\ d_2\le y_2\\
& g(d_1,d_2;t,u_1,u_2,r_1,r_2,v)\le\zeta(d_1+d_2)+\eta(d_1+y_2), && \forall d_1\le y_1,\ y_2\le d_2,\ d_1+d_2\le y_1+y_2\\
& g(d_1,d_2;t,u_1,u_2,r_1,r_2,v)\le\zeta(d_1+d_2)+\eta(y_1+d_2), && \forall d_2\le y_2,\ y_1\le d_1,\ d_1+d_2\le y_1+y_2\\
& g(d_1,d_2;t,u_1,u_2,r_1,r_2,v)\le(\zeta+\eta)(y_1+y_2), && \forall d_1\ge y_1,\ d_2\ge y_2\\
& g(d_1,d_2;t,u_1,u_2,r_1,r_2,v)\le\zeta(y_1+y_2)+\eta(d_1+y_2), && \forall d_1\le y_1,\ y_2\le d_2,\ d_1+d_2\ge y_1+y_2\\
& g(d_1,d_2;t,u_1,u_2,r_1,r_2,v)\le\zeta(y_1+y_2)+\eta(y_1+d_2), && \forall d_2\le y_2,\ y_1\le d_1,\ d_1+d_2\ge y_1+y_2.
\end{aligned} \tag{EC.8}$$

For notational brevity, we will denote the function g by $g(d_1,d_2)$, ignoring the implicit dependence on the dual variables t,u_1,u_2,r_1,r_2 and v. Note that the dual feasible set is the set of all bi-quadratic functions $g(d_1,d_2)$ that are bounded above by a piecewise linear function with six facets (one for each constraint). Let $q_i(d_1,d_2)$ denote the linear function for facet i, i.e., the right hand side of the constraint i in model (EC.8).

Let us consider the case where $g(d_1,d_2)$ touches the piecewise linear function at exactly 6 points, one on each facet. To find these points, for each i, we equate $\nabla g(d_1,d_2)=\nabla q_i(d_1,d_2)$ and solve for (d_1^*,d_2^*) as a function of the dual variables t,u_1,u_2,r_1,r_2,v. Then, setting $g(d_1^*,d_2^*)=q_i(d_1^*,d_2^*)$ gives us a condition on the dual variables for which the two functions touch at exactly one point. We for now ignore the ranges of d_1,d_2 in which each constraint is valid (we will later use these ranges to establish constraints on the dual variables). Table EC.1 gives, for each facet, the points of contact and the condition on dual variables t,u_1,u_2,r_1,r_2,v. Note that we have the following six equations that need to be satisfied for $g(d_1,d_2)$ to touch all six facets of the piecewise linear function:

$$t=\frac{u_1(\eta+\zeta-r_2)^2+u_2(\eta+\zeta-r_1)^2+v((\eta+\zeta)(r_1+r_2)-r_1r_2-(\eta+\zeta)^2)}{4u_1u_2-v^2},$$

Facet i	$\nabla g(d_1^*, d_2^*) = \nabla q_i(d_1^*, d_2^*)$	$g(d_1^*, d_2^*) = q_i(d_1^*, d_2^*)$
1	$(d_1^*, d_2^*) = \left(\frac{(\eta+\zeta)(2u_2-v)-2r_1u_2+r_2v}{4u_1u_2-v^2}, \frac{(\eta+\zeta)(2u_1-v)-2r_2u_1+r_1v}{4u_1u_2-v^2}\right)$	$t = \frac{u_1(\eta+\zeta-r_2)^2+u_2(\eta+\zeta-r_1)^2+v((\eta+\zeta)(r_1+r_2)-r_1r_2-(\eta+\zeta)^2)}{4u_1u_2-v^2}$
2	$(d_1^*, d_2^*) = \left(\frac{2(\eta+\zeta)u_2-\zeta v-2r_1u_2+r_2v}{4u_1u_2-v^2}, \frac{-(\eta+\zeta)v+2\zeta u_1-2r_2u_1+r_1v}{4u_1u_2-v^2}\right)$	$t = \eta y_2 + \frac{u_1(\zeta-r_2)^2+u_2(\eta+\zeta-r_1)^2+v(\zeta r_1+(\eta+\zeta)r_2-r_1r_2-\zeta(\eta+\zeta))}{4u_1u_2-v^2}$
3	$(d_1^*, d_2^*) = \left(\frac{-(\eta+\zeta)v+2\zeta u_2-2r_1u_2+r_2v}{4u_1u_2-v^2}, \frac{2(\eta+\zeta)u_1-\zeta v-2r_2u_1+r_1v}{4u_1u_2-v^2}\right)$	$t = \eta y_1 + \frac{u_2(\zeta-r_1)^2+u_1(\eta+\zeta-r_2)^2+v(\zeta r_2+(\eta+\zeta)r_1-r_1r_2-\zeta(\eta+\zeta))}{4u_1u_2-v^2}$
4	$(d_1^*, d_2^*) = \left(\frac{-2r_1u_2+r_2v}{4u_1u_2-v^2}, \frac{-2r_2u_1+r_1v}{4u_1u_2-v^2}\right)$	$t = (\eta+\zeta)(y_1+y_2) + \frac{r_2^2u_1+r_1^2u_2-r_1r_2v}{4u_1u_2-v^2}$
5	$(d_1^*, d_2^*) = \left(\frac{2\eta u_2-2r_1u_2+r_2v}{4u_1u_2-v^2}, \frac{-\eta v-2r_2u_1+r_1v}{4u_1u_2-v^2}\right)$	$t = \zeta y_1 + (\eta+\zeta)y_2 + \frac{r_2^2u_1+u_2(\eta-r_1)^2+vr_2(\eta-r_1)}{4u_1u_2-v^2}$
6	$(d_1^*, d_2^*) = \left(\frac{-\eta v-2r_1u_2+r_2v}{4u_1u_2-v^2}, \frac{2\eta u_1-2r_2u_1+r_1v}{4u_1u_2-v^2}\right)$	$t = \zeta y_2 + (\eta+\zeta)y_1 + \frac{r_1^2u_2+u_1(\eta-r_2)^2+vr_1(\eta-r_2)}{4u_1u_2-v^2}$

Table EC.1 Points of contact of biquadratic with each facet, and conditions on $(t, u_1, u_2, r_1, r_2, v)$ for biquadratic and facet to touch at exactly one point.

$$
\begin{aligned}
t &= \eta y_2 + \frac{u_1(\zeta-r_2)^2+u_2(\eta+\zeta-r_1)^2+v(\zeta r_1+(\eta+\zeta)r_2-r_1r_2-\zeta(\eta+\zeta))}{4u_1u_2-v^2}, \\
t &= \eta y_1 + \frac{u_2(\zeta-r_1)^2+u_1(\eta+\zeta-r_2)^2+v(\zeta r_2+(\eta+\zeta)r_1-r_1r_2-\zeta(\eta+\zeta))}{4u_1u_2-v^2}, \\
t &= (\eta+\zeta)(y_1+y_2) + \frac{r_2^2u_1+r_1^2u_2-r_1r_2v}{4u_1u_2-v^2}, \\
t &= \zeta y_1 + (\eta+\zeta)y_2 + \frac{r_2^2u_1+u_2(\eta-r_1)^2+vr_2(\eta-r_1)}{4u_1u_2-v^2}, \\
t &= \zeta y_2 + (\eta+\zeta)y_1 + \frac{r_1^2u_2+u_1(\eta-r_2)^2+vr_1(\eta-r_2)}{4u_1u_2-v^2}
\end{aligned}
$$

Thus, we know that the dual variables need to be of this form so that the biquadratic touches all six facets.

In fact, this is an overdetermined system of equations. And we have that for any v, this system of equations is solved by

$$u_1 = \frac{(\eta+\zeta)v}{2\zeta}, \tag{EC.9}$$

$$u_2 = \frac{(\eta+\zeta)v}{2\zeta}, \tag{EC.10}$$

$$r_1 = \frac{\zeta(\eta+\zeta) - 2\left(y_1(\eta+\zeta)+y_2\zeta\right)v}{2\zeta}, \tag{EC.11}$$

$$r_2 = \frac{\zeta(\eta+\zeta) - 2\left(y_1\zeta+y_2(\eta+\zeta)\right)v}{2\zeta}, \tag{EC.12}$$

$$t = \frac{\zeta^2(\eta+\zeta)^2 + 2\zeta(\eta+\zeta)(\eta+2\zeta)(y_1+y_2)v + 2(\eta+2\zeta)\left((\eta+\zeta)(y_1+y_2)^2 - 2\eta y_1y_2\right)v^2}{4\zeta(\eta+2\zeta)v} \tag{EC.13}$$

Substituting these dual variables into the intersection points (d_1^*, d_2^*) of Table EC.1, we get the contact points in Table EC.2. We still need to check whether these contact points occur in the corresponding ranges of (d_1, d_2) of (EC.8). We can easily check that since $\eta > 0$ and $\zeta > 0$, then the dual variables are feasible (i.e., the contact points are in the required range) if $v < 0$ (see Table EC.2).

Thus, we consider the following optimization program:

$$\sup_{v<0} a + bv + \frac{c}{v}$$

where,

$$a = \frac{1}{2}(\eta+\zeta)(y_1+y_2+m_1+m_2),$$

Facet i	**Contact points**	**Condition on v**
1	$(d_1^*,d_2^*)=\left(y_1+\frac{\zeta(\eta+\zeta)}{2(\eta+2\zeta)v},y_2+\frac{\zeta(\eta+\zeta)}{2(\eta+2\zeta)v}\right)$	$d_1^*\le y_1,\ d_2^*\le y_2 \Leftrightarrow v<0$
2	$(d_1^*,d_2^*)=\left(y_1+\frac{\zeta(\eta+3\zeta)}{2(\eta+2\zeta)v},y_2-\frac{\zeta(\eta+\zeta)}{2(\eta+2\zeta)v}\right)$	$d_1^*\le y_1,\ y_2\le d_2^*,\ d_1^*+d_2^*\le y_1+y_2 \Leftrightarrow v<0$
3	$(d_1^*,d_2^*)=\left(y_1-\frac{\zeta(\eta+\zeta)}{2(\eta+2\zeta)v},y_2+\frac{\zeta(\eta+3\zeta)}{2(\eta+2\zeta)v}\right)$	$d_2^*\le y_2,\ y_1\le d_1^*,\ d_1^*+d_2^*\le y_1+y_2 \Leftrightarrow v<0$
4	$(d_1^*,d_2^*)=\left(y_1-\frac{\zeta(\eta+\zeta)}{2(\eta+2\zeta)v},y_2-\frac{\zeta(\eta+\zeta)}{2(\eta+2\zeta)v}\right)$	$d_1^*\ge y_1,\ d_2^*\ge y_2 \Leftrightarrow v<0$
5	$(d_1^*,d_2^*)=\left(y_1+\frac{\zeta(\eta+\zeta)}{2(\eta+2\zeta)v},y_2-\frac{\zeta(\eta+3\zeta)}{2(\eta+2\zeta)v}\right)$	$d_1^*\le y_1,\ y_2\le d_2^*,\ d_1^*+d_2^*\ge y_1+y_2 \Leftrightarrow v<0$
6	$(d_1^*,d_2^*)=\left(y_1-\frac{\zeta(\eta+3\zeta)}{2(\eta+2\zeta)v},y_2+\frac{\zeta(\eta+\zeta)}{2(\eta+2\zeta)v}\right)$	$d_2^*\le y_2,\ y_1\le d_1^*,\ d_1^*+d_2^*\ge y_1+y_2 \Leftrightarrow v<0$

Table EC.2 Condition on v so that the points of contact of biquadratic with each facet occurs in the required range.

$$b=\frac{1}{2\zeta}\left[(\eta+\zeta)\left((m_1-y_1+m_2-y_2)^2+\sigma_1^2+\sigma_2^2\right)+2\zeta\rho\sigma_1\sigma_2-2\eta(m_1-y_1)(m_2-y_2)\right],$$
$$c=\frac{\zeta(\eta+\zeta)^2}{4(\eta+2\zeta)},$$

where $b>0$ since we have that $\zeta>0$, $\eta>0$, $\sigma_1,\sigma_2>0$, and $\rho\in(-1,1)$. Note that the objective function is the objective of a dual feasible solution (EC.9)–(EC.13) parametrized by v. The supremum is achieved at

$$v^*=-\sqrt{\frac{c}{b}}=\frac{-\zeta(\eta+\zeta)}{\sqrt{2(\eta+2\zeta)\left[(\eta+\zeta)\left((m_1-y_1+m_2-y_2)^2+\sigma_1^2+\sigma_2^2\right)+2\zeta\rho\sigma_1\sigma_2-2\eta(m_1-y_1)(m_2-y_2)\right]}}. \tag{EC.14}$$

The optimal value is

$$a+bv^*+\frac{c}{v^*}=\frac{1}{2}(\eta+\zeta)(y_1+y_2+m_1+m_2)-(\eta+\zeta)\Phi\left(\mathbf{y};\frac{\eta+3\zeta}{\eta+\zeta}\right), \tag{EC.15}$$

where recall $\Phi(\cdot)$ is defined in (8). We know that (EC.15) is the dual objective of a dual feasible solution to (EC.7), though we have not shown that it is the dual optimal value. By weak duality, (EC.15) is a *lower bound* on the primal optimal value $M(y_1,y_2)$ defined in (EC.6). This proves the first part of Proposition 1.

To prove the second part of Proposition 1, we need to show that under Condition 1 strong duality holds and that (EC.15) is the dual *optimal* value. To do this, we need to construct a feasible demand distribution whose primal value matches the right-hand side of (EC.15). That is, we will show that there exists a six-point distribution $f_y^*\in\mathcal{F}$ whose primal objective value is

$$\mathbb{E}_{f_y^*}\left[\zeta\min(\tilde{d}_1+\tilde{d}_2,y_1+y_2)+\sum_{j=1,2}\eta\min(\tilde{d}_j,y_j)\right]=\frac{1}{2}(\eta+\zeta)(y_1+y_2+m_1+m_2)-(\eta+\zeta)\Phi\left(\mathbf{y};\frac{\eta+3\zeta}{\eta+\zeta}\right).$$

To construct the distribution, we use the contact points of the biquadratic to each facet as the support points (i.e., the support points in Table EC.2) with the optimal v^* defined in (EC.14). Note that the support points of this distribution are summarized in Table EC.3, where

$$\nu:=\frac{(\eta+3\zeta)}{(\eta+\zeta)}$$

Facet i	**Support Point**	**Probability**
1	$(d_1^*, d_2^*) = (y_1 - \Phi(\mathbf{y}),\ y_2 - \Phi(\mathbf{y}))$	π_1
2	$(d_1^*, d_2^*) = (y_1 - \nu\Phi(\mathbf{y}),\ y_2 + \Phi(\mathbf{y}))$	π_2
3	$(d_1^*, d_2^*) = (y_1 + \Phi(\mathbf{y}),\ y_2 - \nu\Phi(\mathbf{y}))$	π_3
4	$(d_1^*, d_2^*) = (y_1 + \Phi(y),\ y_2 + \Phi(\mathbf{y}))$	π_4
5	$(d_1^*, d_2^*) = (y_1 - \Phi(\mathbf{y}),\ y_2 + \nu\Phi(\mathbf{y}))$	π_5
6	$(d_1^*, d_2^*) = (y_1 + \nu\Phi(\mathbf{y}),\ y_2 - \Phi(\mathbf{y}))$	π_6

Table EC.3 The support points and the corresponding probabilities in a worst-case probability distribution $f^*_{y,\pi}$, where (EC.16)–(EC.20) hold.

In order for this demand distribution to be feasible, the probabilities $\pi_1, \ldots, \pi_6$ must satisfy the following conditions on the moments:

$$\begin{aligned}
\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 + \pi_6 &= 1 \\
-\pi_1 - \nu\pi_2 + \pi_3 + \pi_4 - \pi_5 + \nu\pi_6 &= \frac{m_1 - y_1}{\Phi(\mathbf{y})}, \\
-\pi_1 + \pi_2 - \nu\pi_3 + \pi_4 + \nu\pi_5 - \pi_6 &= \frac{m_2 - y_2}{\Phi(\mathbf{y})}, \\
2y_1\left(-\pi_1 - \nu\pi_2 + \pi_3 + \pi_4 - \pi_5 + \nu\pi_6\right) + \Phi(\mathbf{y})\left(\pi_1 + \nu^2\pi_2 + \pi_3 + \pi_4 + \pi_5 + \nu^2\pi_6\right) &= \frac{m_1^2 + \sigma_1^2 - y_1^2}{\Phi(\mathbf{y})} \\
2y_2\left(-\pi_1 + \pi_2 - \nu\pi_3 + \pi_4 + \nu\pi_5 - \pi_6\right) + \Phi(\mathbf{y})\left(\pi_1 + \pi_2 + \nu^2\pi_3 + \pi_4 + \nu^2\pi_5 + \pi_6\right) &= \frac{m_2^2 + \sigma_2^2 - y_2^2}{\Phi(\mathbf{y})} \\
y_1\left(-\pi_1 + \pi_2 - \nu\pi_3 + \pi_4 + \nu\pi_5 - \pi_6\right) + y_2\left(-\pi_1 - \nu\pi_2 + \pi_3 + \pi_4 - \pi_5 + \nu\pi_6\right) & \\
+\Phi(\mathbf{y})\left(\pi_1 - \nu\pi_2 - \nu\pi_3 + \pi_4 - \nu\pi_5 - \nu\pi_6\right) &= \frac{m_1 m_2 + \rho\sigma_1\sigma_2 - y_1 y_2}{\Phi(\mathbf{y})}
\end{aligned}$$

This is a system of six equations with six unknowns. However, this is an overdetermined system of equations, since for any π_6, we can check that the probabilities $\pi_1, \ldots, \pi_5$ defined below solve the equations:

$$\begin{aligned}
\pi_1 &= \frac{2\zeta}{(\eta+\zeta)}\pi_6 + \frac{(\eta+\zeta)((m_2-y_2)^2+\sigma_2^2) + (\eta+3\zeta)((m_1-y_1)(m_2-y_2)+\rho\sigma_1\sigma_2)}{4(\eta+2\zeta)\Phi(\mathbf{y})^2} \\
&\quad - \frac{(\eta+3\zeta)(m_1-y_1) + (\eta+\zeta)(m_2-y_2)}{4(\eta+2\zeta)\Phi(\mathbf{y})} && \text{(EC.16)} \\
\pi_2 &= -\pi_6 + \frac{(\eta+\zeta)^2}{4\zeta(\eta+2\zeta)}\left(\frac{(m_1-y_1)^2+\sigma_1^2-\Phi(\mathbf{y})^2}{\Phi(\mathbf{y})^2}\right) && \text{(EC.17)} \\
\pi_3 &= -\pi_6 + \frac{(\eta+\zeta)^2\left((m_1-y_1)^2+(m_2-y_2)^2+\sigma_1^2+\sigma_2^2-2\Phi(\mathbf{y})^2\right)}{8\zeta(\eta+2\zeta)\Phi(\mathbf{y})^2} + \frac{(\eta+\zeta)(m_1-y_1-m_2+y_2)}{4(\eta+2\zeta)\Phi(\mathbf{y})} && \text{(EC.18)} \\
\pi_4 &= -\frac{2\zeta}{(\eta+\zeta)}\pi_6 - \frac{(\eta-\zeta)(\eta+\zeta)\left((m_1-y_1)^2+\sigma_1^2\right)}{8\zeta(\eta+2\zeta)\Phi(\mathbf{y})^2} - \frac{(\eta+\zeta)^2\left((m_2-y_2)^2+\sigma_2^2\right)}{8\zeta(\eta+2\zeta)\Phi(\mathbf{y})^2} \\
&\quad + \frac{(\eta-\zeta)(\eta+\zeta)}{8\zeta(\eta+2\zeta)} + \frac{(\eta+3\zeta)^2}{8\zeta(\eta+2\zeta)} + \frac{(\eta+\zeta)(m_2-y_2)}{4(\eta+2\zeta)\Phi(\mathbf{y})} + \frac{(\eta+3\zeta)(m_1-y_1)}{4(\eta+2\zeta)\Phi(\mathbf{y})} && \text{(EC.19)} \\
\pi_5 &= \pi_6 - \frac{(\eta+\zeta)^2\left((m_1-y_1)^2+\sigma_1^2-(m_2-y_2)^2-\sigma_2^2\right)}{8\zeta(\eta+2\zeta)\Phi(\mathbf{y})^2} - \frac{(\eta+\zeta)(m_1-y_1-m_2+y_2)}{4(\eta+2\zeta)\Phi(\mathbf{y})} && \text{(EC.20)}
\end{aligned}$$

We can check that for any probability distribution $(\pi_1, \ldots, \pi_6)$ satisfying (EC.16)–(EC.20), the expected cost is

$$\begin{aligned}
&\mathbb{E}\left[\eta\min(d_1,y_1) + \eta\min(d_2,y_2) + \zeta\min(d_1+d_2,y_1+y_2)\right] \\
&\quad = \pi_1\left[\eta(y_1-\Phi(\mathbf{y})) + \eta(y_2-\Phi(\mathbf{y})) + \zeta(y_1+y_2-2\Phi(\mathbf{y}))\right]
\end{aligned}$$

$$
\begin{aligned}
&+\pi_2\left[\eta(y_1-\nu\Phi(\mathbf{y}))+\eta y_2+\zeta(y_1+y_2+(1-\zeta)\Phi(\mathbf{y}))\right]\\
&+\pi_3\left[\eta y_1+\eta(y_2-\nu\Phi(\mathbf{y}))+\zeta(y_1+y_2+(1-\nu)\Phi(\mathbf{y}))\right]\\
&+\pi_4\left[\eta y_1+\eta y_2+\zeta(y_1+y_2)\right]\\
&+\pi_5\left[\eta(y_1-\Phi(\mathbf{y}))+\eta y_2+\zeta(y_1+y_2)\right]\\
&+\pi_6\left[\eta y_1+\eta(y_2-\Phi(\mathbf{y}))+\zeta(y_1+y_2)\right]\\
=&(\eta+\zeta)(y_1+y_2)(\pi_1+\ldots+\pi_6)-\Phi(\mathbf{y})\left[2(\eta+\zeta)\pi_1+(\eta+2\zeta)\pi_2+(\eta+2\zeta)\pi_3+\eta\pi_5+\eta\pi_6\right]\\
=&\frac{1}{2}(\eta+\zeta)\left(y_1+y_2+m_1+m_2\right)-(\eta+\zeta)\Phi(\mathbf{y}).
\end{aligned}
$$

Therefore, as long as the probabilities $\pi_1,\ldots,\pi_6$ are nonnegative, we have found a primal feasible demand distribution whose expected cost matches (EC.15), i.e., the optimal value of a dual feasible solution. Hence, this proves that

$$M(y_1,y_2)=\frac{1}{2}(\eta+\zeta)\left(y_1+y_2+m_1+m_2\right)-(\eta+\zeta)\Phi(\mathbf{y}). \tag{EC.21}$$

Note that, based on (EC.16)–(EC.20), $\pi_1,\ldots,\pi_6$ are nonnegative if and only if $\max(0,\alpha_1,\alpha_5)\le\pi_6\le\min(\alpha_2,\alpha_3,\alpha_4)$, where:

$$
\begin{aligned}
\alpha_1:=&\frac{(\eta+\zeta)}{2\zeta}\Bigg(\frac{(\eta+3\zeta)(m_1-y_1)+(\eta+\zeta)(m_2-y_2)}{4(\eta+2\zeta)\Phi(\mathbf{y})}\\
&\quad-\frac{(\eta+\zeta)((m_2-y_2)^2+\sigma_2^2)+(\eta+3\zeta)((m_1-y_1)(m_2-y_2)+\rho\sigma_1\sigma_2)}{4(\eta+2\zeta)\Phi(\mathbf{y})^2}\Bigg)\\
\alpha_2:=&\frac{(\eta+\zeta)^2}{4\zeta(\eta+2\zeta)}\left(\frac{(m_1-y_1)^2+\sigma_1^2-\Phi(\mathbf{y})^2}{\Phi(\mathbf{y})^2}\right)\\
\alpha_3:=&\frac{(\eta+\zeta)^2\left((m_1-y_1)^2+(m_2-y_2)^2+\sigma_1^2+\sigma_2^2-2\Phi(\mathbf{y})^2\right)}{8\zeta(\eta+2\zeta)\Phi(\mathbf{y})^2}+\frac{(\eta+\zeta)(m_1-y_1-m_2+y_2)}{4(\eta+2\zeta)\Phi(\mathbf{y})}\\
\alpha_4:=&\frac{(\eta+\zeta)}{2\zeta}\Bigg(-\frac{(\eta-\zeta)(\eta+\zeta)\left((m_1-y_1)^2+\sigma_1^2\right)}{8\zeta(\eta+2\zeta)\Phi(\mathbf{y})^2}-\frac{(\eta+\zeta)^2\left((m_2-y_2)^2+\sigma_2^2\right)}{8\zeta(\eta+2\zeta)\Phi(\mathbf{y})^2}\\
&\quad+\frac{(\eta-\zeta)(\eta+\zeta)}{8\zeta(\eta+2\zeta)}+\frac{(\eta+3\zeta)^2}{8\zeta(\eta+2\zeta)}+\frac{(\eta+\zeta)(m_2-y_2)}{4(\eta+2\zeta)\Phi(\mathbf{y})}+\frac{(\eta+3\zeta)(m_1-y_1)}{4(\eta+2\zeta)\Phi(\mathbf{y})}\Bigg)\\
\alpha_5:=&\frac{(\eta+\zeta)^2\left((m_1-y_1)^2+\sigma_1^2-(m_2-y_2)^2-\sigma_2^2\right)}{8\zeta(\eta+2\zeta)\Phi(\mathbf{y})^2}+\frac{(\eta+\zeta)(m_1-y_1-m_2+y_2)}{4(\eta+2\zeta)\Phi(\mathbf{y})}
\end{aligned}
$$

Note that these are the same $\alpha_1,\cdots,\alpha_4$ defined in Condition 1, except that we use the simplification $\nu=\frac{\eta+3\zeta}{\eta+\zeta}$, which implies $\nu-1=\frac{2\zeta}{\eta+\zeta}$, $\nu+1=\frac{2(\eta+2\zeta)}{\eta+\zeta}$, $\frac{2-\nu}{\nu-1}=\frac{\eta-\zeta}{2\zeta}$, to rewrite $\alpha_1,\cdots,\alpha_4$ in terms of ν (instead of ζ,η). Hence, under Condition 1, the range $[\max(0,\alpha_1,\alpha_5),\min(\alpha_2,\alpha_3,\alpha_4)]$ for π_6 is nonempty, so there exists a primal feasible solution whose probabilities are nonnegative and where (EC.16)–(EC.20) hold. Hence, the bound is tight under Condition 1. ■

EC.2.4. Proof of Lemma 4

Given the definition of $\alpha_1,\ldots,\alpha_4$ in Condition 1, our goal is to show that if $y_1-m_1=y_2-m_2$ and if (11) holds, then the interval $[\max(0,\alpha_1,\alpha_5),\min(\alpha_2,\alpha_3,\alpha_4)]$ is non-empty.

Assumption EC.1.

$$(\nu-1)\Phi(\mathbf{y})\,(m_2-y_2-m_1+y_1)\geq(m_1-y_1)^2-(m_2-y_2)^2+\sigma_1^2-\sigma_2^2. \tag{EC.22}$$

Note that we can make this assumption *without loss of generality*. This is because, if it does not hold, we simply have to switch indices 1 and 2.

Note that under Assumption EC.1, we have that $\alpha_5\leq 0$ and $\alpha_3\geq\alpha_2$. Hence, under Assumption EC.1, we only need to show that $[\max(0,\alpha_1),\min(\alpha_2,\alpha_4)]$ is non-empty. This is true if $\alpha_2\geq 0$, $\alpha_2\geq\alpha_1$, $\alpha_4\geq 0$, and $\alpha_4\geq\alpha_1$, or equivalently,

$$\frac{(m_1-y_1)^2+\sigma_1^2}{\Phi(\mathbf{y})^2}\geq 1 \tag{EC.23}$$

$$\nu+\frac{(m_1-y_1)^2+\sigma_1^2+\rho\sigma_1\sigma_2+(m_1-y_1)(m_2-y_2)}{\Phi(\mathbf{y})^2}\geq\frac{(m_2-y_2)+\nu(m_1-y_1)}{\Phi(\mathbf{y})} \tag{EC.24}$$

$$\nu+\frac{(m_1-y_1)^2+\sigma_1^2+\rho\sigma_1\sigma_2+(m_1-y_1)(m_2-y_2)}{\Phi(\mathbf{y})^2}\geq 1-\frac{(m_2-y_2)+\nu(m_1-y_1)}{\Phi(\mathbf{y})} \tag{EC.25}$$

$$\nu+\frac{\rho\sigma_1\sigma_2+(m_1-y_1)(m_2-y_2)}{\Phi(\mathbf{y})^2}\geq 0. \tag{EC.26}$$

When we have $m_1-y_1=m_2-y_2=\delta$, then we have $\Phi(\mathbf{y})=\sqrt{\delta^2+\tilde{\gamma}}$, where $\tilde{\gamma}:=\frac{\sigma_1^2+\sigma_2^2+(\nu-1)\rho\sigma_1\sigma_2}{\nu+1}$. Further, condition (EC.22) reduces simply to $\sigma_1\leq\sigma_2$, and the inequalities (EC.23)–(EC.26) simplify to:

$$\tilde{\gamma}\leq\sigma_1^2 \tag{EC.27}$$

$$(\nu+2)\delta^2-(\nu+1)\delta\sqrt{\delta^2+\tilde{\gamma}}\geq\frac{\sigma_1^2+\sigma_2^2-(\nu-1)\sigma_1^2-\tilde{\gamma}(\nu^2+1)}{\nu-1} \tag{EC.28}$$

$$\delta^2+\delta\sqrt{\delta^2+\tilde{\gamma}}\geq\frac{\sigma_1^2+\sigma_2^2-(\nu-1)\sigma_1^2-\tilde{\gamma}(\nu^2-\nu+2)}{\nu^2-1} \tag{EC.29}$$

$$\delta^2\geq\frac{\sigma_1^2+\sigma_2^2-\tilde{\gamma}(\nu^2+1)}{\nu^2-1} \tag{EC.30}$$

We can show that the left hand sides of inequalities (EC.28), (EC.29) and (EC.30) are convex functions in δ minimized at $\delta=0$. Hence, sufficient conditions for (EC.28), (EC.29) and (EC.30) to be satisfied are simply that the right hand sides are less than zero. This is true if $\tilde{\gamma}(\nu^2+1)\geq\sigma_1^2+\sigma_2^2$. Note that this condition immediately implies that the right-hand side of (EC.30) is negative. Further, since the right-hand side of (EC.28) is less than the right-hand side of (EC.29), then we only need to check that $\tilde{\gamma}(\nu^2+1)\geq\sigma_1^2+\sigma_2^2$ implies that the right-hand side of (EC.29) is negative. Indeed:

$$\begin{aligned}\sigma_1^2+\sigma_2^2-(\nu-1)\sigma_1^2-\tilde{\gamma}(\nu^2-\nu+2)&=\sigma_1^2+\sigma_2^2-\tilde{\gamma}(\nu^2+1)-(\nu-1)\sigma_1^2+\tilde{\gamma}(\nu-1)\\&\leq-(\nu-1)\sigma_1^2+\tilde{\gamma}(\nu-1)=(\nu-1)(\tilde{\gamma}-\sigma_1^2)\leq 0.\end{aligned}$$

The first inequality uses $\tilde{\gamma}(\nu^2+1)\geq\sigma_1^2+\sigma_2^2$, and the final inequality uses $\tilde{\gamma}\leq\sigma_1^2$ in (EC.27).

Hence, $\tilde{\gamma}\leq\sigma_1^2$ and $\tilde{\gamma}(\nu^2+1)\geq\sigma_1^2+\sigma_2^2$ are sufficient conditions to ensure that the interval $[\max(0,\alpha_1,\alpha_5),\min(\alpha_2,\alpha_3,\alpha_4)]$ is non-empty if $m_1-y_1=m_2-y_2$ and $\sigma_1\leq\sigma_2$. Note that these sufficient conditions are equivalent to (11). ■

EC.2.5. Proof of Theorem 1

From Lemma 1 of the paper, we know that

$$\mathbb{E}_f\left[C(\mathbf{y},\mathbf{D})\right] = h(y_1+y_2-m_1-m_2)+s_0(m_1+m_2)+(s-s_0)\mathbb{E}_f\left[(d_1-y_1)^+ + (d_1-y_1)^+ + \zeta(d_1+d_2-y_1-y_2)^+\right],$$

where $\zeta = \frac{p+h-s}{s-s_0}$. Therefore, the result follows directly from Proposition 1. ■

EC.2.6. Proof of Proposition 2

Consider the function

$$\bar{C}(\mathbf{y}) = s_0(m_1+m_2) - \frac{1}{2}(p-h-s_0)(y_1+y_2-m_1-m_2) + (p+h-s_0)\Phi\left(\mathbf{y}; \tfrac{3(p+h-s)+s-s_0}{p+h-s_0}\right),$$

where $\bar{C}(\mathbf{y})$ is the upper bound defined in (10). We note from the definition of $\Phi(\mathbf{y})$ in (8) that for $i=1,2$:

$$\frac{\partial}{\partial y_i}\Phi\left(\mathbf{y}; \tfrac{3(p+h-s)+s-s_0}{p+h-s_0}\right) = -\frac{(p+h-s_0)(m_i-y_i)+(p+h-s)(m_{-i}-y_{-i})}{2\left(2(p+h)-s-s_0\right)\Phi(\mathbf{y})}.$$

Hence, we have that

$$\frac{\partial \bar{C}}{\partial y_1} = -\frac{1}{2}(p-h-s_0) + (p+h-s_0)\left[\frac{(p+h-s_0)(y_1-m_1)+(p+h-s)(y_2-m_2)}{2(2(p+h)-s-s_0)\Phi(\mathbf{y})}\right] \tag{EC.31}$$

$$\frac{\partial \bar{C}}{\partial y_2} = -\frac{1}{2}(p-h-s_0) + (p+h-s_0)\left[\frac{(p+h-s)(y_1-m_1)+(p+h-s_0)(y_2-m_2)}{2(2(p+h)-s-s_0)\Phi(\mathbf{y})}\right]. \tag{EC.32}$$

We can easily check function $\bar{C}$ is jointly convex in (y_1,y_2). Hence, its global minimum is achieved at $(\bar{y}_1,\bar{y}_2)$ where $\nabla\bar{C}(\bar{y}_1,\bar{y}_2)=0$. For (EC.31)–(EC.31) to be simultaneously zero, it must be that $\bar{y}_1-m_1=\bar{y}_2-m_2$. Let us define $\delta := y_1-m_1 = y_2-m_2$ and find the value of δ where the partial derivatives (EC.31)–(EC.31) are zero. Note that if $y_1-m_1=y_2-m_2=\delta$, then

$$\Phi(\mathbf{y}) = \sqrt{\delta^2 + \frac{(p+h-s_0)(\sigma_1^2+\sigma_2^2)+2(p+h-s)\rho\sigma_1\sigma_2}{2(2(p+h)-s-s_0)}},$$

hence

$$\frac{\partial \bar{C}}{\partial y_1} = \frac{\partial \bar{C}}{\partial y_2} = -\frac{1}{2}(p-h-s_0) + \frac{1}{2}(p+h-s_0)\delta \Bigg/ \sqrt{\delta^2 + \frac{(p+h-s_0)(\sigma_1^2+\sigma_2^2)+2(p+h-s)\rho\sigma_1\sigma_2}{2(2(p+h)-s-s_0)}}. \tag{EC.33}$$

Setting (EC.33) to zero and solving for δ, we have

$$\delta = \frac{1}{2}(p-h-s_0)\sqrt{\frac{(p+h-s_0)(\sigma_1^2+\sigma_2^2)+2(p+h-s)\rho\sigma_1\sigma_2}{2h(p-s_0)(2(p+h)-s-s_0)}}.$$

Hence, we have that $\bar{C}$ is minimized at $\bar{\mathbf{y}}=(\bar{y}_1,\bar{y}_2)$ where

$$\bar{y}_1-m_1=\bar{y}_2-m_2=\frac{1}{2}(p-h-s_0)\sqrt{\frac{(p+h-s_0)(\sigma_1^2+\sigma_2^2)+2(p+h-s)\rho\sigma_1\sigma_2}{2h(p-s_0)(2(p+h)-s-s_0)}}.$$

Note that

$$\min_{\mathbf{y}}\bar{C}(\mathbf{y}) = \bar{C}(\bar{\mathbf{y}}) = s_0(m_1+m_2) + \sqrt{\frac{2h(p-s_0)\left[(p+h-s_0)(\sigma_1^2+\sigma_2^2)+2(p+h-s)\rho\sigma_1\sigma_2\right]}{2(p+h)-s-s_0}}. \tag{EC.34}$$

To complete the proof, we also need to show that $\bar{\mathbf{y}}$ also is the minimizer of $C(\mathbf{y}) := \sup_{f\in F}\mathbb{E}_f\left[C(\mathbf{y},\tilde{\mathbf{D}})\right]$. Note that, since (σ_1,σ_2,ρ) satisfy condition (11) and since $\bar{y}_1-m_1=\bar{y}_2-m_2$, then we know from Lemma 4 that Condition 1 is true. Hence, by Theorem 1 it follows that $C(\bar{\mathbf{y}}) = \bar{C}(\bar{\mathbf{y}})$. Furthermore, since $C(\mathbf{y})$ is jointly convex in $\mathbf{y} = (y_1, y_2)$, we have that for any $\mathbf{y}$,

$$C(\mathbf{y}) \geq C(\bar{\mathbf{y}}) + \nabla C(\bar{\mathbf{y}})^\top (\mathbf{y} - \bar{\mathbf{y}}) = \bar{C}(\bar{\mathbf{y}}). \tag{EC.35}$$

Hence, $C(\mathbf{y}) \geq C(\bar{\mathbf{y}})$ for all $\mathbf{y}$. This proves that $\bar{\mathbf{y}}$ is the minimizer of C, and

$$C^* := \min_{\mathbf{y}} C(\mathbf{y}) = C(\bar{\mathbf{y}}) = \bar{C}(\bar{\mathbf{y}}) = s_0(m_1 + m_2) + \sqrt{\frac{2h(p-s_0)\left[(p+h-s_0)(\sigma_1^2+\sigma_2^2) + 2(p+h-s)\rho\sigma_1\sigma_2\right]}{2(p+h)-s-s_0}},$$

where the last equality is from (EC.34). ■

EC.3. Bound on the Benefit of Pooling

If there are two locations (same setting as Section 4), then we can check that the difference in the expected cost of an unpooled system and a pooled system is

$$B_f(y_1, y_2) = (p+h-s) \cdot \mathbb{E}_f \left[(\tilde{d}_1 - y_1)^+ + (\tilde{d}_2 - y_2)^+ - (\tilde{d}_1 + \tilde{d}_2 - y_1 - y_2)^+\right]. \tag{EC.36}$$

In other words, it is proportional to the reduction in the unmet demand by pooling inventory. Note that the benefit of pooling depends on the demand distribution.

Given only mean and covariance information, the largest possible benefit of pooling is found as

$$\sup_{f\in\mathcal{F}} B_f(y_1,y_2) = (p+h-s)\cdot \sup_{f\in\mathcal{F}} \mathbb{E}_f \left[(\tilde{d} - y_1)^+ + (\tilde{d}_2 - y_2)^+ - (\tilde{d}_1 + \tilde{d}_2 - y_1 - y_2)^+\right],$$

where $\mathcal{F}$ consists of joint distributions of $(\tilde{d}_1, \tilde{d}_2)$ with mean (m_1, m_2), variance (σ_1^2, σ_2^2), and correlation coefficient ρ.

The demands are dependent, so we define the following events:

$$\begin{aligned}
I_1 &: d_1 \geq y_1,\ d_2 \geq y_2,\ d_1 + d_2 - y_1 - y_2 > 0,\\
I_2 &: d_1 < y_1,\ d_2 \geq y_2,\ d_1 + d_2 - y_1 - y_2 \geq 0,\\
I_3 &: d_1 < y_1,\ d_2 \geq y_2,\ d_1 + d_2 - y_1 - y_2 < 0,\\
I_4 &: d_1 \geq y_1,\ d_2 < y_2,\ d_1 + d_2 - y_1 - y_2 \geq 0,\\
I_5 &: d_1 \geq y_1,\ d_2 < y_2,\ d_1 + d_2 - y_1 - y_2 < 0,\\
I_6 &: d_1 < y_1,\ d_2 < y_2,\ d_1 + d_2 - y_1 - y_2 < 0,
\end{aligned}$$

Define the following variables, for $i = 1, \ldots, 6$

$$\begin{pmatrix} p(i) \\ D_1(i) \\ D_2(i) \\ D_{11}(i) \\ D_{22}(i) \\ D_{12}(i) \end{pmatrix} = \begin{pmatrix} \Pr(I_i) \\ \mathbb{E}(\tilde{d}_1|I_i)\Pr(I_i) \\ \mathbb{E}(\tilde{d}_2|I_i)\Pr(I_i) \\ \mathbb{E}(\tilde{d}_1^2|I_i)\Pr(I_i) \\ \mathbb{E}(\tilde{d}_2^2|I_i)\Pr(I_i) \\ \mathbb{E}(\tilde{d}_1\tilde{d}_2|I_i)\Pr(I_i) \end{pmatrix}$$

Therefore, note that

$$
\begin{aligned}
&\mathbb{E}_f\left[(\tilde{d}-y_1)^+ + (\tilde{d}_2-y_2)^+ - (\tilde{d}_1+\tilde{d}_2-y_1-y_2)^+\right]\\
&=\mathbb{E}(\tilde{d}_1-y_1|I_1)\Pr(I_1)+\mathbb{E}(\tilde{d}_1-y_1|I_4)\Pr(I_4)+\mathbb{E}(\tilde{d}_1-y_1|I_5)\Pr(I_5)\\
&\quad+\mathbb{E}(\tilde{d}_2-y_2|I_1)\Pr(I_1)+\mathbb{E}(\tilde{d}_2-y_2|I_2)\Pr(I_2)+\mathbb{E}(\tilde{d}_2-y_2|I_3)\Pr(I_3)\\
&\quad-\mathbb{E}(\tilde{d}_1+\tilde{d}_2-y_1-y_2|I_1)\Pr(I_1)-\mathbb{E}(\tilde{d}_1+\tilde{d}_2-y_1-y_2|I_2)\Pr(I_2)-\mathbb{E}(\tilde{d}_1+\tilde{d}_2-y_1-y_2|I_4)\Pr(I_4)\\
&=D_1(1)+D_1(4)+D_1(5)-y_1p(1)-y_1p(4)-y_1p(5)\\
&\quad+D_2(1)+D_2(2)+D_2(3)-y_2p(1)-y_2p(2)-y_2p(3)\\
&\quad-D_1(1)-D_2(1)-D_1(2)-D_2(2)-D_1(4)-D_2(4)\\
&\quad+(y_1+y_2)\times(p(1)+p(2)+p(4))\\
&=y_1\times p(2)-D_1(2)-y_2p(3)+D_2(3)+y_2\times p(4)-D_2(4)-y_1p(5)+D_1(5)
\end{aligned}
$$

If we define matrices

$$
\begin{aligned}
\mathbf{M}_i &:= \begin{pmatrix} p(i) & D_1(i) & D_2(i)\\ D_1(i) & D_{11}(i) & D_{12}(i)\\ D_2(i) & D_{12}(i) & D_{22}(i)\end{pmatrix}, \quad i=1,2,3,4,5,6,\\
\mathbf{Q}_2 &:= \begin{pmatrix} y_1 & -\frac{1}{2} & 0\\ -\frac{1}{2} & 0 & 0\\ 0 & 0 & 0\end{pmatrix}, \quad \mathbf{Q}_3 := \begin{pmatrix} -y_2 & 0 & \frac{1}{2}\\ 0 & 0 & 0\\ \frac{1}{2} & 0 & 0\end{pmatrix},\\
\mathbf{Q}_4 &:= \begin{pmatrix} y_2 & 0 & -\frac{1}{2}\\ 0 & 0 & 0\\ -\frac{1}{2} & 0 & 0\end{pmatrix}, \quad \mathbf{Q}_5 := \begin{pmatrix} -y_1 & \frac{1}{2} & 0\\ \frac{1}{2} & 0 & 0\\ 0 & 0 & 0\end{pmatrix}
\end{aligned}
$$

then the largest benefit of pooling can be found by solving

$$
\begin{aligned}
\sup_{f\in\mathcal{F}} B_f(y_1,y_2) = \ &\underset{\mathbf{M}_i}{\text{maximize}} && \operatorname{Tr}\left(\mathbf{Q}_2(y_1)\cdot\mathbf{M}_2+\mathbf{Q}_3(y_2)\cdot\mathbf{M}_3+\mathbf{Q}_4(y_2)\cdot\mathbf{M}_4+\mathbf{Q}_5(y_1)\cdot\mathbf{M}_5\right)\\
&\text{subject to} && \sum_{i=1}^{6}\mathbf{M}_i = \begin{pmatrix} 1 & m_1 & m_2\\ m_1 & \sigma_1^2+m_1^2 & \rho\sigma_1\sigma_2+m_1m_2\\ m_2 & \rho\sigma_1\sigma_2+m_1m_2 & \sigma_2^2+m_2^2\end{pmatrix} := \mathbf{\Sigma}_0,\\
& && \mathbf{M}_1,\mathbf{M}_2,\ldots,\mathbf{M}_6 \succeq 0
\end{aligned}
$$

EC.4. Asymmetric Costs in the Two-Location Model

Consider a two-location model where the fulfillment costs are asymmetric. Specifically, s_{0i} is the in-location fulfillment cost in location i, and the cross-location cost s_{12} is not necessarily equal to s_{21}. The cost function is given by:

$$
\begin{aligned}
C(\mathbf{y},\mathbf{D}) = &\ h\cdot(y_1+y_2-d_1-d_2)^+ + p\cdot(d_1+d_2-y_1-y_2)^+\\
&+s_{01}\cdot\min(y_1,d_1)+s_{02}\cdot\min(y_2,d_2)\\
&+s_{12}\cdot\underbrace{\min\left((y_1-d_1)^+,(d_2-y_2)^+\right)}_{\text{quantity shipped from location 1 to location 2}}\\
&+s_{21}\cdot\underbrace{\min\left((y_2-d_2)^+,(d_1-y_1)^+\right)}_{\text{quantity shipped from location 2 to location 1}}
\end{aligned}
\tag{EC.37}
$$

Using the identities $\min(a,b) = a - (a-b)^+$ and $a^+ = a + (-a)^+$, we have the following simplified cost function:

$$\begin{aligned} C(\mathbf{y},\mathbf{D}) = {} & h\cdot(y_1+y_2) + (s_{01}-h)d_1 + (s_{02}-h)d_2 + (p+h)\cdot(d_1+d_2-y_1-y_2)^+ \\ & - s_{01}(d_1-y_1)^+ - s_{02}(d_2-y_2)^+ + s_{12}\cdot\min\Big((y_1-d_1)^+, (d_2-y_2)^+\Big) \\ & + s_{21}\cdot\min\left((d_1-y_1)^+, (y_2-d_2)^+\right) \end{aligned} \tag{EC.38}$$

We next describe how to numerically solve the problem with asymmetric costs. The demands are dependent, so we define the following events:

$$\begin{aligned} &I_1 : d_1 \geq y_1,\ d_2 \geq y_2,\ d_1+d_2-y_1-y_2 \geq 0, \\ &I_2 : d_1 < y_1,\ d_2 \geq y_2,\ d_1+d_2-y_1-y_2 \geq 0, \\ &I_3 : d_1 < y_1,\ d_2 \geq y_2,\ d_1+d_2-y_1-y_2 < 0, \\ &I_4 : d_1 \geq y_1,\ d_2 < y_2,\ d_1+d_2-y_1-y_2 \geq 0, \\ &I_5 : d_1 \geq y_1,\ d_2 < y_2,\ d_1+d_2-y_1-y_2 < 0, \\ &I_6 : d_1 < y_1,\ d_2 < y_2,\ d_1+d_2-y_1-y_2 < 0, \end{aligned}$$

Define the following variables, for $i = 1,\ldots,6$

$$\begin{pmatrix} p(i) \\ D_1(i) \\ D_2(i) \\ D_{11}(i) \\ D_{22}(i) \\ D_{12}(i) \end{pmatrix} = \begin{pmatrix} \Pr(I_i) \\ \mathbb{E}(\tilde{d}_1|I_i)\Pr(I_i) \\ \mathbb{E}(\tilde{d}_2|I_i)\Pr(I_i) \\ \mathbb{E}(\tilde{d}_1^2|I_i)\Pr(I_i) \\ \mathbb{E}(\tilde{d}_2^2|I_i)\Pr(I_i) \\ \mathbb{E}(\tilde{d}_1\tilde{d}_2|I_i)\Pr(I_i) \end{pmatrix}$$

Note that

$$\begin{aligned} \mathbb{E}(d_1+d_2-y_1-y_2)^+ &= \sum_{i=1,2,4} \mathbb{E}(d_1+d_2-y_1-y_2|I_i)\Pr(I_i) \\ \mathbb{E}(d_1-y_1)^+ &= \sum_{i=1,4,5} \mathbb{E}(d_1-y_1|I_i)\Pr(I_i) \\ \mathbb{E}(d_2-y_2)^+ &= \sum_{i=1,2,3} \mathbb{E}(d_2-y_2|I_i)\Pr(I_i) \\ \mathbb{E}\min\Big((y_1-d_1)^+, (d_2-y_2)^+\Big) &= \mathbb{E}(y_1-d_1|I_2)\Pr(I_2) + \mathbb{E}(d_2-y_2|I_3)\Pr(I_3) \\ \mathbb{E}\min\Big((d_1-y_1)^+, (y_2-d_2)^+\Big) &= \mathbb{E}(y_2-d_2|I_4)\Pr(I_4) + \mathbb{E}(d_1-y_1|I_5)\Pr(I_5) \end{aligned}$$

Therefore, from (EC.38), we have

$$\begin{aligned} \mathbb{E}[C(\mathbf{y},\mathbf{D})] = {} & h(y_1+y_2) + (s_{01}-h)m_1 + (s_{02}-h)m_2 \\ & + (p+h-s_{01})\cdot D_1(1) + (p+h-s_{02})\cdot D_2(1) - ((p+h-s_{01})y_1 + (p+h-s_{02})y_2)\cdot p(1) \\ & + (p+h-s_{12})\cdot D_1(2) + (p+h-s_{02})\cdot D_2(2) - ((p+h-s_{12})y_1 + (p+h-s_{02})y_2)\cdot p(2) \\ & + 0\cdot D_1(3) + (s_{12}-s_{02})\cdot D_2(3) - (s_{12}-s_{02})y_2\cdot p(3) \\ & + (p+h-s_{01})\cdot D_1(4) + (p+h-s_{21})\cdot D_2(4) - ((p+h-s_{01})y_1 + (p+h-s_{21})y_2)\cdot p(4) \\ & + (s_{21}-s_{01})\cdot D_1(5) + 0\cdot D_2(5) - (s_{21}-s_{01})y_1\cdot p(5) \end{aligned}$$

Defining $\eta_1 = p+h-s_{01}$, $\eta_2 = p+h-s_{02}$, $\zeta_{12} = p+h-s_{12}$, $\zeta_{21} = p+h-s_{21}$. Note that $s_{12}-s_{02} = \eta_2-\zeta_{12}$ and $s_{21}-s_{01} = \eta_1-\zeta_{21}$. Therefore,

$$\begin{aligned}\mathbb{E}[C(\mathbf{y},\mathbf{D})] = h(y_1+y_2) &+ (s_{01}-h)m_1 + (s_{02}-h)m_2 \\ &+ \eta_1\cdot D_1(1) + \eta_2\cdot D_2(1) - (\eta_1 y_1 + \eta_2 y_2)\cdot p(1) \\ &+ \zeta_{12}\cdot D_1(2) + \eta_2\cdot D_2(2) - (\zeta_{12} y_1 + \eta_2 y_2)\cdot p(2) \\ &+ 0\cdot D_1(3) + (\eta_2-\zeta_{12})\cdot D_2(3) - (\eta_2-\zeta_{12})y_2\cdot p(3) \\ &+ \eta_1\cdot D_1(4) + \zeta_{21}\cdot D_2(4) - (\eta_1 y_1 + \zeta_{21} y_2)\cdot p(4) \\ &+ (\eta_1-\zeta_{21})\cdot D_1(5) + 0\cdot D_2(5) - (\eta_1-\zeta_{21})y_1\cdot p(5)\end{aligned}$$

If we define matrices

$$\begin{aligned}
\mathbf{M}_i &:= \begin{pmatrix} p(i) & D_1(i) & D_2(i) \\ D_1(i) & D_{11}(i) & D_{12}(i) \\ D_2(i) & D_{12}(i) & D_{22}(i) \end{pmatrix}, \quad i=1,2,3,4,5,6, \\
\mathbf{Q}_1 &:= \begin{pmatrix} -\eta_1 y_1 - \eta_2 y_2 & \frac{\eta_1}{2} & \frac{\eta_2}{2} \\ \frac{\eta_1}{2} & 0 & 0 \\ \frac{\eta_2}{2} & 0 & 0 \end{pmatrix}, \\
\mathbf{Q}_2 &:= \begin{pmatrix} -\zeta_{12} y_1 - \eta_2 y_2 & \frac{\zeta_{12}}{2} & \frac{\eta_2}{2} \\ \frac{\zeta_{12}}{2} & 0 & 0 \\ \frac{\eta_2}{2} & 0 & 0 \end{pmatrix}, \quad \mathbf{Q}_3 := \begin{pmatrix} -(\eta_2-\zeta_{12})y_2 & 0 & \frac{\eta_2-\zeta_{12}}{2} \\ 0 & 0 & 0 \\ \frac{\eta_2+\zeta_{12}}{2} & 0 & 0 \end{pmatrix}, \\
\mathbf{Q}_4 &:= \begin{pmatrix} -\eta_1 y_1 - \zeta_{21} y_2 & \frac{\eta_1}{2} & \frac{\zeta_{21}}{2} \\ \frac{\eta_1}{2} & 0 & 0 \\ \frac{\zeta_{21}}{2} & 0 & 0 \end{pmatrix}, \quad \mathbf{Q}_5 := \begin{pmatrix} -(\eta_1-\zeta_{21})y_1 & \frac{\eta_1-\zeta_{21}}{2} & 0 \\ \frac{\eta_1-\zeta_{21}}{2} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}
\end{aligned}$$

then the worst-case expected cost can be found for a given $\mathbf{y}$ by solving

$$\begin{aligned}
\sup_{f\in\mathcal{F}} \mathbb{E}_f\left[C(\mathbf{y},\tilde{\mathbf{D}})\right] = \underset{\mathbf{M}_i}{\text{maximize}} \quad & h(y_1+y_2) + (s_0-h)(m_1+m_2) + \text{Tr}\left(\sum_{i=1}^{5} \mathbf{Q}_i(\mathbf{y})\cdot\mathbf{M}_i\right) \\
\text{subject to} \quad & \sum_{i=1}^{6} \mathbf{M}_i = \begin{pmatrix} 1 & m_1 & m_2 \\ m_1 & \sigma_1^2+m_1^2 & \rho\sigma_1\sigma_2+m_1m_2 \\ m_2 & \rho\sigma_1\sigma_2+m_1m_2 & \sigma_2^2+m_2^2 \end{pmatrix} := \mathbf{\Sigma}_0, \\
& \mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_6 \succeq 0
\end{aligned}$$

Therefore, we know that we can solve the two-location problem with asymmetric costs via a SDP. To see how such a numerical approach should extend to the case where there are more than two locations, note that events $I_1,\ldots,I_6$ enumerates all of the possible outcomes on which locations have excess/deficit and the degree of imbalance. The number of outcomes would determine the number of matrices involved in the SDP constraint. However, the number of outcomes is also exponential in the number of locations. Therefore, even a numerical approach is not tractable beyond just a few locations.

Nevertheless, we can use the case of two-location asymmetric costs to demonstrate how we can address a single location, 2-period robust inventory problem (see Figure EC.2). Consider a single location problem where y_i is the inventory replenishment in period $i=1,2$. We can set $y_2=0$ if this is an initial inventory problem without replenishment. Each period faces a random demand d_1, d_2, and unfulfilled demand is lost. The fulfillment cost is s_0 in any period. Unused inventory in period 1 carries over to the next period at a cost

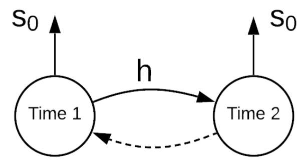


Figure EC.2 Two period model

h. At the end of both periods, any unmet demand incurs a penalty of c_u, and any excess inventory incurs a cost c_o.

This single location, 2-period problem can be thought of as 2-location, single period problem with asymmetric costs. The total costs can be written as

$$\begin{aligned}
C(\mathbf{y},\mathbf{D}) =& s_0\min(d_1,y_1)+h(y_1-d_1)^+ + s_0\min(d_2,y_2)+s_0\underbrace{\min\left((y_1-d_1)^+,(d_2-y_2)^+\right)}_{\substack{\text{unused inventory in period 1}\\ \text{used for period 2 fulfillment}}} \\
&\underbrace{+c_o(y_1+y_2-d_1-d_2)^+ + c_u(d_1+d_2-y_1-y_2)^+}_{\text{underage and overage cost if period 2 inventory could be used for period 1}} \\
&\underbrace{+(c_o+c_u)\min\left((y_2-d_2)^+,(d_1-y_1)^+\right)}_{\text{additional cost since period 2 inventory could not be used for period 1 demand}} \\
=& c_o(y_1+y_2-d_1-d_2)^+ + c_u(d_1+d_2-y_1-y_2)^+ \\
&+hy_1+(s_0-h)\min(d_1,y_1)+s_0\min(d_2,y_2) \\
&+s_0\min\left((y_1-d_1)^+,(d_2-y_2)^+\right)+(c_o+c_u)\min\left((y_2-d_2)^+,(d_1-y_1)^+\right)
\end{aligned}$$

Comparing this last expression to (EC.37), we can define $s_{01}=s_0-h$, $s_{02}=s_0$, $p=c_u$, $h=c_o$, $s_{12}=s_0$, $s_{21}=c_o+c_u$, and then solve the worst-case expected 2-period cost as a SDP.

We solved an instance of the 2-period robust inventory level problem for $c_u=9$, $c_o=2$, $h=1$, $s_0=3$, $m_1=m_2=100$, $\sigma_1=\sigma_2=50$, and for differing values of the correlation ρ between the demands in period 1 and period 2. We set $y_2=0$ and solve for the optimal y_1 that minimizes the worst-case expected cost. Note that the SDP can only be solved for a given value of (y_1,y_2). Hence, to solve for the optimal y_1, we perform Golden-section search. Figure EC.3 shows the robust initial inventory level as plotted against the correlation coefficient ρ.

EC.5. Experiments with a 2-level Nested Cost Structure

We study the performance of our proposed heuristic in a 2-level nested cost structure, i.e., where fulfillment costs for spillover demands are constant ($s_{ij}=s>s_0$ for all $i\neq j$). The simple 2-level structure is useful because we can isolate the effect of pooling from the cost structure.

In these experiments, $n=5$, $h=2$, $s=1$, $s_0=0.5$, and we vary $p\in\{0.5,1,2,4,8\}$. We randomly generate 50 instances of the joint distribution. We do this by first randomly generating the parameters of the marginal distribution of demand in each of the $n=5$ locations. (We run experiments on different families of marginal distributions: Normal, exponential, beta prime, and Student-t). Then, we generate a random correlation matrix based on Numpacharoen and Atsawarungruangkit (2012). A Gaussian copula is used to describe the dependence between the random variables.

Below we describe how we randomly generate the parameters of the marginal distributions.

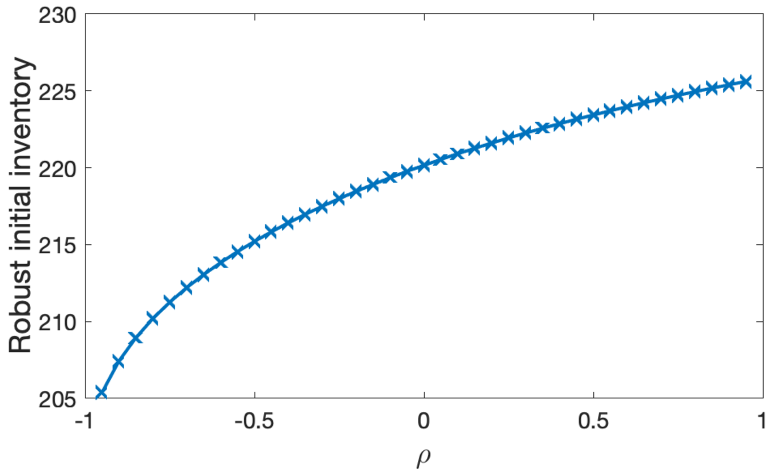


Figure EC.3 Optimal robust initial inventory for the 2-period model

1. Normal: the means are identical with $m = 300$, and the standard deviation is chosen at random from $[100, 800]$.
2. Exponential: the mean of the exponential distribution is chosen at random from [100,500]. The standard deviation is equal to the mean.
3. BetaPrime: the mean is fixed at $m = 2$. The parameters α and β are chosen as follows. β is chosen at random from $[2, 3]$, and $\alpha = m \cdot (\beta - 1)$.
4. Student-t: the parameter ν is chosen at random from $[2, 3]$.

Given a specific joint demand distribution f, we estimate the optimal expected cost $C_f^* := \min_{\mathbf{y}\geq 0} \mathbb{E}_f \left[C(\mathbf{y}, \tilde{\mathbf{D}})\right]$ using sample average approximation with 10^4 samples of the demand vector. Using only the mean and the covariance of the random demand, we use our heuristic to approximate the robust inventory levels with $\mathbf{y}^{\mathrm{H}}$. We then compute the expected cost of the heuristic solution under the known true distribution f (which we denote as C_f^{H}).

Figure EC.4 illustrates the gap $C_f^{\mathrm{H}} - C_f^*$ (the $\circ$ markers), which represents the performance of the heuristic under a specific distribution f. We observe that the performance of the heuristic depends on p, seen from the small optimality gap for small values of p. If the distribution is either Normal or exponential, the heuristic has an actual expected cost that is close to optimal even for high values of p, with relative gaps in the order of .1% or 1%. For the beta prime and the Student-t distribution families, we observe that the relative gap can be as high as the order of 10%.

The figure also shows the gap $C^{\mathrm{H}} - C_f^*$ (the $+$ markers). Since under beta prime or Student-t distributions, the circle markers are close to the plus markers, we can infer that the expected cost under these distributions is close to the worst-case expected cost in the neighborhood of $\mathbf{y}^{\mathrm{H}}$. (Note that the plus markers are always above the circle markers since $C_f^{\mathrm{H}} \leq C^{\mathrm{H}}$.) We next discuss how the performance of a distributionally robust heuristic can be improved for these cases.

Since there are multiple joint ditributions in $\mathcal{F}$, then the range of possible values of $\mathbb{E}_f[C(\mathbf{y}, \tilde{\mathbf{D}})]$ for a given $\mathbf{y}$ could potentially be wide. This ambiguity may result in the true optimal solution to be different from the

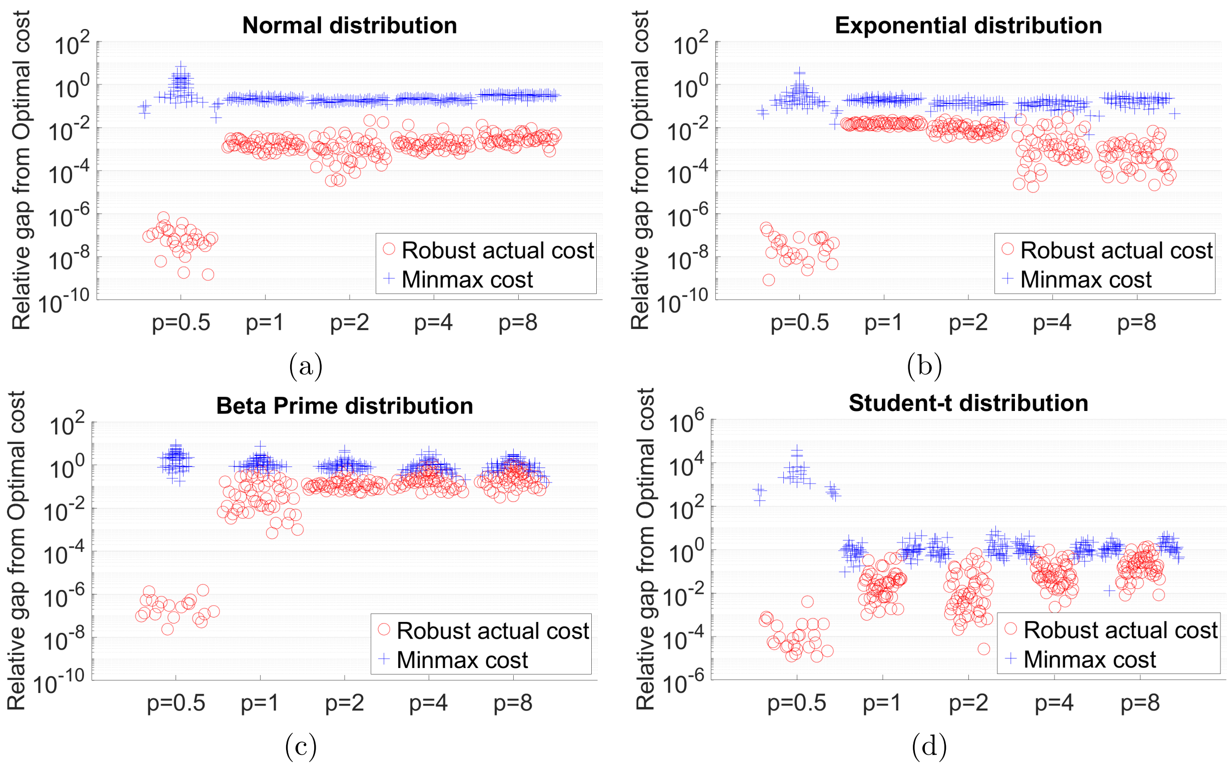


Figure EC.4 **Relative gaps $(C_f^{\mathbf{H}} - C_f^*)/C_f^*$ (robust actual cost) and $(C^{\mathbf{H}} - C^*)/C^*$ (minmax cost). The x-axis spread of the data around each value of p is solely for visual clarity.**

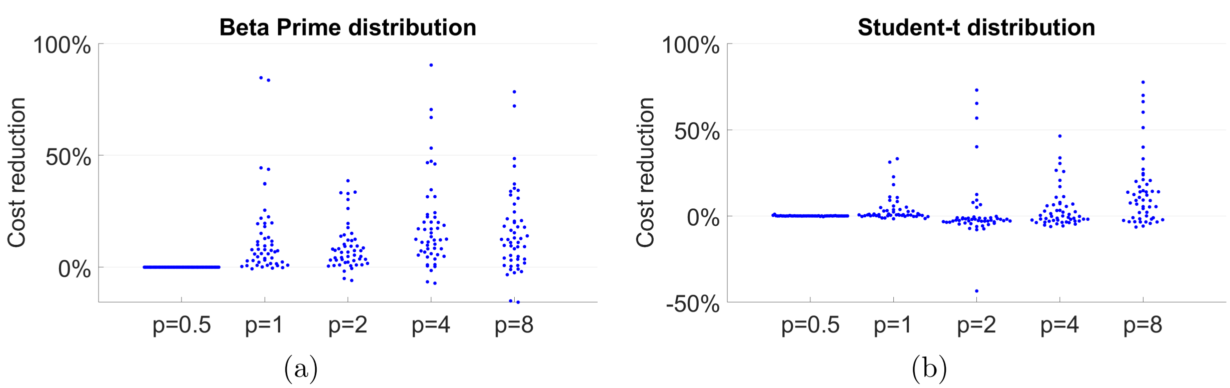


Figure EC.5 **Reduction in expected cost (under the true distribution) of the robust inventory levels with partitioned statistics information.**

robust solution under some distributions (e.g. under beta prime or Student-t). A way to reduce ambiguity is by further restricting the distribution set, which can be accomplished by adding more information to $\mathcal{F}$. This can be done with *partitioned statistics* information, specifically, the mean and covariance of random vector $(\tilde{\mathbf{D}}^+, \tilde{\mathbf{D}}^-)$ whose i^{th} elements are $(\tilde{d}_i - m_i)^+$ and $(m_i - \tilde{d}_i)^+$, respectively. Partitioned statistics measures asymmetry of the distribution that is not represented by covariance alone (Natarajan et al. 2017). Moreover, we can utilize the techniques from this paper, hence adapt Proposition 3, for a distributionally robust heuristic under this additional information (see Section EC.6 in the Appendix for the complete formulation). Figure EC.5 shows that the additional information could significantly reduce the expected cost of the distributionally robust inventory levels. It is no surprise that asymmetry information is important to estimate the impact of pooling, as this is in line with results from the pooling literature, specifically Yang and Schrage (2009) who show that right-skewed demand distributions can cause inventory levels to increase rather than decrease under pooling.

EC.6. Asymmetry Information

Based on Natarajan et al. (2017), we incorporate into our robust models the partitioned statistics information. Specifically, the mean and covariance of random vector $(\tilde{\mathbf{D}}^+, \tilde{\mathbf{D}}^-)$ whose i^{th} elements are $(\tilde{d}_i - m_i)^+$ and $(m_i - \tilde{d}_i)^+$, respectively, are defined to be:

$$\mathbb{E}\left[\begin{pmatrix}\tilde{\mathbf{D}}^+\\ \tilde{\mathbf{D}}^-\end{pmatrix}\right] =: \bar{\mathbf{m}} \qquad \mathbb{E}\left[\begin{pmatrix}\tilde{\mathbf{D}}^+\\ \tilde{\mathbf{D}}^-\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{D}}^+\\ \tilde{\mathbf{D}}^-\end{pmatrix}^\intercal\right] =: \bar{\mathbf{Q}} \tag{EC.39}$$

The set of distributions that the random demand can take is defined as $\bar{\mathcal{F}}_{\geq 0}$, which specifies that the random demand has non-negative support, with mean $\mathbf{m}$, and with mean and covariance of the partitioned statistics given in (EC.39). We follow the same approach as in Theorem 4.3 in Natarajan et al. (2017) to adapt Proposition 3 to derive the following upper bound including the partitioned statistics information. We omit the proof to avoid repetition.

PROPOSITION EC.1. *For the n-location newsvendor problem under inventory risk pooling with a L-level nested fulfillment cost structure, we have* $\sup_{f\in\bar{\mathcal{F}}_{\geq 0}} \mathbb{E}_f[C(\mathbf{y},\tilde{\mathbf{D}})] \leq \bar{\bar{C}}_L(\mathbf{y})$ *for any* $\mathbf{y}\in\Re^n$*, where*

$$\begin{aligned}
\bar{\bar{C}}_L(\mathbf{y}) := \min_{\substack{t_0,\mathbf{t},\mathbf{Y},\mathbf{u},\\ \mathbf{B},\mathbf{W},\mathbf{U},\mathbf{V}}} \quad & h\cdot\mathbf{e}^\top(\mathbf{y}-\mathbf{m}) + \mathbf{s}_0^\top\mathbf{m} + t_0 + \mathbf{t}^\top\bar{\mathbf{m}} + \langle \mathbf{Y},\bar{\mathbf{Q}}\rangle + \mathbf{e}^\top\mathbf{B}\mathbf{e} \\
\text{s.t.} \quad & \begin{pmatrix} t_0 & \frac{1}{2}\mathbf{t}^\top & \frac{1}{2}\mathbf{u}^\top \\ \frac{1}{2}\mathbf{t} & \mathbf{Y} & -\frac{1}{2}\mathbf{V}^\top \\ \frac{1}{2}\mathbf{u} & -\frac{1}{2}\mathbf{V} & \mathbf{U}\end{pmatrix} \succeq 0, \\
& \mathbf{u} = -\mathbf{W}\mathbf{e} + \left(\mathbf{B}+\mathbf{B}^\top\right)\mathbf{e} + \mathbf{P}(\mathbf{y}-\mathbf{m}), \\
& \mathbf{V} \geq \bar{\mathbf{P}}, \\
& \mathbf{U} \leq \mathbf{W} - \mathbf{B}, \\
& \mathbf{W},\mathbf{B} \geq 0, \\
& t_0\in\Re,\ \mathbf{t}\in\Re^{2n},\ \mathbf{u}\in\Re^N,\ \mathbf{Y}\in\Re^{2n\times 2n},\ \mathbf{B},\mathbf{W},\mathbf{U}\in\Re^{N\times N},\ \mathbf{V}\in\Re^{N\times 2n},
\end{aligned} \tag{EC.40}$$

with $\mathbf{P} := \left(\mathbf{E}_{L-1}^\top \mathrm{diag}(\eta_{L-1})\ \mathbf{E}_{L-2}^\top \mathrm{diag}(\eta_{L-2})\ \cdots\ \mathbf{E}_0^\top \mathrm{diag}(\eta_0)\right)^\top \in \Re^{N\times n}$*, and* $\bar{\mathbf{P}} = \begin{bmatrix}\mathbf{P} & -\mathbf{P}\end{bmatrix} \in \Re^{N\times 2n}$.

The heuristic solution can be similarly obtained by setting $\mathbf{y}$ as a decision variable, constrained by $\mathbf{y}\geq 0$.

EC.7. Multiple Demand Channels

To simplify our discussion, we consider a two-level nested fulfillment cost structure for the online demand (i.e., where fulfillment cost for spillover demand is constant), though the technique can be generalized to an L-level structure. Let p_b and p_o be the penalty cost of unmet brick-and-mortar store demand and online demand, respectively. The per-unit overage cost is h. We normalize the cost for meeting store demand to zero. As before, the cost of in-location fulfillment of online demand is s_0, and the cost of spillover demand fulfillment is s, where $s > s_0$. For a customer region $j \in [n]$, let $\tilde{d}_j^o$ and $\tilde{d}_j^b$ be the stochastic online demand and the stochastic store demand, respectively. We denote the vector of online demands as $\tilde{\mathbf{D}}^o = (\tilde{d}_j^o)$ and the vector of store demands as $\tilde{\mathbf{D}}^b = (\tilde{d}_j^b)$. We let $\tilde{\mathbf{D}} = (\tilde{\mathbf{D}}^b, \tilde{\mathbf{D}}^o)$ as the vector of all demands with a mean vector $\mathbf{m} = (\mathbf{m}^b, \mathbf{m}^o)$ and covariance matrix $\mathbf{\Sigma}$.

Store demand can only be met with inventory from the same location. However, online demand can be fulfilled from inventory from any location. We assume that $p_o + h > s$, that $p_b + h > s$, and that $p_b + s_0 > p_o$. Given our assumptions on the cost parameters, it is optimal for each local store to first meet the store demands to the maximum extent possible, then for excess inventory to be used to fulfill in-location online demand, before demand spillover is fulfilled. To see why, note that since $p_b + s_0 > p_o$, then it is cheaper to use an inventory unit to meet store demand than to fulfill a local online demand. Moreover, the assumptions imply that $p_b + h + s_0 > s$, so it is cheaper to allow an online demand to spillover than to use in-location fulfillment and not meet a store demand. Therefore, we can write the cost as

$$\begin{aligned}
C(\mathbf{y}, \mathbf{D}) = h \cdot \left(\sum_{j\in[n]} \left(y_j - d_j^b\right)^+ - \sum_{j\in[n]} d_j^o \right)^+ + p_o \cdot \left(\sum_{j\in[n]} d_j^o - \sum_{j\in[n]} \left(y_j - d_j^b\right)^+ \right)^+ \\
+ p_b \cdot \sum_{j\in[n]} \left(d_j^b - y_j\right)^+ + s_0 \cdot \left(\sum_{j\in[n]} d_j^o - \sum_{j\in[n]} \left(d_j^o - \left(y_j - d_j^b\right)^+ \right)^+ \right) \\
+ s \cdot \left(\sum_{j\in[n]} \left(d_j^o - \left(y_j - d_j^b\right)^+ \right)^+ - \left(\sum_{j\in[n]} d_j^o - \sum_{j\in[n]} \left(y_j - d_j^b\right)^+ \right)^+ \right)
\end{aligned}$$

We observe that, due to the presence of store demand which is prioritized due to its lower cost of fulfillment, the cost structure is more complicated than before. In particular, the last term in the cost function has a composition of a function $f(x) = (a - x)^+$ and $g(\mathbf{x}) = \sum_j x_j^+$. This requires a careful treatment in developing the tractable SDP heuristic. We first simplify the cost function by reducing the number of such terms using the relationship that if $a \geq 0$, then $(a - (b - c)^+) = (a + c - b)^+ - (c - b)^+$. Also using the fact that $(c - b)^+ = b - c + (c - b)^+$, we can simplify the cost function to

$$\begin{aligned}
C(\mathbf{y}, \mathbf{D}) = h \cdot \mathbf{e}^\top \left(\mathbf{y} - \mathbf{D}^o - \mathbf{D}^b\right) + s_0 \cdot \mathbf{e}^\top \mathbf{D}^o + (h + p_b + s_o - s) \cdot \sum_{j\in[n]} \left(d_j^b - y_j\right)^+ \\
+ (s - s_0) \cdot \sum_{j\in[n]} \left(d_j^o + d_j^b - y_j\right)^+ + (h + p_o - s) \cdot \left(\sum_{j\in[n]} \left(d_j^o + d_j^b - y_j\right) - \sum_{j\in[n]} \left(d_j^b - y_j\right)^+ \right)^+
\end{aligned}$$

We define the constants $\gamma := h + p_b + s_o - s$, $\eta_0 := s - s_0$, and $\eta_1 := h + p_o - s$. Hence, the minmax expected cost under the omni-channel demand is equivalent to

$$C_{\text{o}}^* := \min_{\mathbf{y}} \left((s_0 - h) \cdot \mathbf{e}^\top \mathbf{m}_o + h \cdot \mathbf{e}^\top \left(\mathbf{y} - \mathbf{m}_s\right) + M_{\text{o}}(\mathbf{y}) \right)$$

where $M_{\text{o}}(\mathbf{y})$ is the optimal value of the moment problem

$$\begin{aligned}
M_{\text{o}}(\mathbf{y}) := \max_{f\in\mathcal{F}} \quad \mathbb{E}_f \Bigg[\gamma \cdot \sum_{j\in[n]} \left(\tilde{d}_j^b - y_j\right)^+ + \eta_0 \cdot \sum_{j\in[n]} \left(\tilde{d}_j^o + \tilde{d}_j^b - y_j\right)^+ \\
+ \eta_1 \cdot \left(\sum_{j\in[n]} \left(\tilde{d}_j^o + \tilde{d}_j^b - y_j\right) - \sum_{j\in[n]} \left(\tilde{d}_j^b - y_j\right)^+ \right)^+ \Bigg].
\end{aligned}$$

We can write the moment problem as

$$M_{\mathrm{o}}(\mathbf{y}) = \max_{f\in\mathcal{F}_{\geq 0}} \quad \mathbb{E}_f \Bigg[\max_{\mathbf{x}^{(0)}\in\{0,1\}^n, x^{(1)}\in\{0,1\}, \mathbf{z}\in\{0,1\}^n} \gamma \cdot \mathbf{z}^\top \left(\tilde{\mathbf{D}}^b - \mathbf{y}\right) + \eta_0 \cdot \mathbf{x}^{(0)\top} \left(\tilde{\mathbf{D}}^o + \tilde{\mathbf{D}}^b - \mathbf{y}\right), \\ + \eta_1 \cdot x^{(1)} \cdot \left(\mathbf{e}^\top \left(\tilde{\mathbf{D}}^o + \tilde{\mathbf{D}}^b - \mathbf{y}\right) - \mathbf{z}^\top \left(\tilde{\mathbf{D}}^b - \mathbf{y}\right)\right) \Bigg]$$

To see why, note that the coefficient of z_j is equal to $\left(\gamma - \eta_1 x^{(1)}\right) \cdot \left(\tilde{d}_j^b - y_j\right)$. Based on our assumptions on the cost parameters, we have that $\gamma = h + p_b + s_o - s > 0$, and $\gamma - \eta_1 = p_b - p_o + s_0 > 0$. Therefore, z_j is equal to 1 if and only if $d_j^b - y_j \geq 0$. Note that unlike in the previous section where the newly introduced variables only interact with other constants or the random demand, we have cross interactions between the new variables from the term $x^{(1)} \cdot \mathbf{z}$. Hence, we introduce a new n-dimensional vector $\mathbf{w} = x^{(1)} \cdot \mathbf{z}$.

Consider the $(3n+1)$-dimensional random vector $\tilde{\mathbf{x}} := \left(\tilde{x}^{(1)\top} \; \tilde{\mathbf{x}}^{(0)\top} \; \tilde{\mathbf{z}}^\top \; \tilde{\mathbf{w}}\right)^\top$, which collects all the new binary variables into a single vector. We again have the following transformation

$$\begin{aligned} \mathbf{x} &:= \mathbb{E}_f\left(\tilde{\mathbf{x}}\right) \in \Re^{3n+1}, \\ \mathbf{Q} &:= \mathbb{E}_f\left(\tilde{\mathbf{x}}\tilde{\mathbf{D}}^\top\right) \in \Re^{(3n+1)\times(2n)}, \\ \mathbf{R} &:= \mathbb{E}_f\left(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\top\right) \in \Re^{(3n+1)\times(3n+1)}. \end{aligned}$$

Therefore, we have linearized the objective to

$$\gamma \cdot \sum_{j\in[n]} \left(Q_{1+n+j,j} - x_{1+n+j} \cdot y_j\right) + \eta_0 \cdot \sum_{j\in[n]} \left(Q_{1+j,j} + Q_{1+j,n+j} - x_{1+j} \cdot y_j\right) \\ + \eta_1 \cdot \sum_{j\in[n]} \left(Q_{1,j} + Q_{1,n+j} - x_1 \cdot y_j - Q_{2n+1+j,j} + x_{2n+1+j} \cdot y_j\right)$$

The constraints are the same as before, but with the addition of a few other constraints that follow from the fact that $\tilde{w}_j = \tilde{x}^{(1)} \cdot \tilde{z}_j$ for all $j \in [n]$. In particular, note that

$$\begin{aligned} R_{1,n+1+j} &= x_{1+2n+j}, && \forall j \in [n], \\ R_{1+n+i,1+2n+j} &= R_{1+2n+1,1+2n+j}, && \forall i \in [n], j \in [n]. \end{aligned}$$

The first constraint follows since the left-hand side is by definition equal to $\mathbb{E}_f\left(\tilde{x}^{(1)} \cdot \tilde{z}_j\right)$, and the right-hand side is $\mathbb{E}_f\left(\tilde{w}_j\right)$. In the second constraint, the left-hand side is equal to $\mathbb{E}_f\left(\tilde{z}_i \cdot \tilde{w}_j\right) = \mathbb{E}_f\left(\tilde{x}^{(1)}\tilde{z}_i\tilde{z}_j\right)$. The right-hand side is equal to $\mathbb{E}_f\left(\tilde{w}_i \cdot \tilde{w}_j\right) = \mathbb{E}_f\left(\tilde{x}^{(1)}\tilde{x}_i\tilde{x}_j\right)$ since $(\tilde{x}^{(1)})^2 = \tilde{x}^{(1)}$. Due to the nonnegativity of demand, aside from the constraint that $\mathbf{Q} \geq 0$, we also have that $\mathbf{z} \leq \mathbf{x}^{(0)}$. This is because $d_j^b - y_j \geq 0$ impies that $d_j^b + d_j^o - y_j \geq 0$, which is equivalent to the condition that $\tilde{z}_j \leq \tilde{x}_j^{(0)}$.

PROPOSITION EC.2. *For the n-location newsvendor problem under inventory risk pooling with online and store demand in each location, if the fulfillment costs of spillover online demand are all equal to s, then* $\sup_{f\in\mathcal{F}_{\geq 0}} \mathbb{E}_f[C(\mathbf{y},\tilde{\mathbf{D}})] \leq \bar{C}_o(\mathbf{y})$ *for any* $\mathbf{y}\in\Re^n$*, where*

$$\begin{aligned}
\bar{C}_o(\mathbf{y}) := \min_{\substack{t_0,\mathbf{t},\mathbf{Y},\mathbf{u},\\ \mathbf{B},\mathbf{W},\mathbf{U},\mathbf{V},\\ \mathbf{g},\mathbf{h},\mathbf{H}}} \quad & h\cdot\mathbf{e}^\top\left(\mathbf{y}-\mathbf{m}^o-\mathbf{m}^b\right)+s_0\cdot\mathbf{e}^\top\mathbf{m}^o+t_0+\mathbf{t}^\top\mathbf{m}+\langle\mathbf{Y},\boldsymbol{\Sigma}+\mathbf{m}\mathbf{m}^\top\rangle+\mathbf{e}^\top\mathbf{B}\mathbf{e}\\
\text{s.t.}\quad & \begin{pmatrix} t_0 & \frac{1}{2}\mathbf{t}^\top & \frac{1}{2}\mathbf{u}^\top\\ \frac{1}{2}\mathbf{t} & \mathbf{Y} & -\frac{1}{2}\mathbf{V}^\top\\ \frac{1}{2}\mathbf{u} & -\frac{1}{2}\mathbf{V} & \mathbf{U}\end{pmatrix}\succeq 0,\\
& \mathbf{u}=-\mathbf{W}\mathbf{e}+\left(\mathbf{B}+\mathbf{B}^\top\right)\mathbf{e}+\begin{pmatrix}\mathbf{0}_{1,n}\\ -\mathbf{I}_n\\ \mathbf{I}_n\\ \mathbf{0}_{n,n}\end{pmatrix}\mathbf{g}+\begin{pmatrix}\mathbf{0}_{1,n}\\ \mathbf{0}_{n,n}\\ \mathbf{0}_{n,n}\\ -\mathbf{I}_n\end{pmatrix}\mathbf{h}+\begin{pmatrix}\eta_1\cdot\mathbf{e}_n^\top\\ \eta_0\cdot\mathbf{I}_n\\ \gamma\cdot\mathbf{I}_n\\ -\eta_0\cdot\mathbf{I}_n\end{pmatrix}\mathbf{y},\\
& \mathbf{V}\geq\begin{pmatrix}\eta_1\cdot\mathbf{e}_n^\top & \eta_1\cdot\mathbf{e}_n^\top\\ \eta_0\cdot\mathbf{I}_n & \eta_0\cdot\mathbf{I}_n\\ \gamma\cdot\mathbf{I}_n & \mathbf{0}_{n,n}\\ -\eta_0\cdot\mathbf{I}_n & \mathbf{0}_{n,n}\end{pmatrix},\\
& \mathbf{U}\leq\mathbf{W}-\mathbf{B}+\begin{pmatrix}\mathbf{0}_{1,n+1} & \mathbf{h}^\top & \mathbf{0}_n^\top\\ \mathbf{0}_{n,n+1} & \mathbf{0}_{n,n} & \mathbf{0}_{n,n}\\ \mathbf{0}_{n,n+1} & \mathbf{0}_{n,n} & \mathbf{H}\\ \mathbf{0}_{n,n+1} & \mathbf{0}_{n,n} & -\mathbf{H}\end{pmatrix},\\
& \mathbf{g},\mathbf{W},\mathbf{B}\geq 0,\\
& t_0\in\Re,\ \mathbf{g},\mathbf{h}\in\Re^n,\ \mathbf{t}\in\Re^{2n},\ \mathbf{u}\in\Re^{3n+1},\ \mathbf{H}\in\Re^{n\times n},\\
& \mathbf{Y}\in\Re^{2n\times 2n},\ \mathbf{B},\mathbf{W},\mathbf{U}\in\Re^{(3n+1)\times(3n+1)},\ \mathbf{V}\in\Re^{(3n+1)\times 2n}.
\end{aligned}\tag{EC.41}$$

Proof. Suppose that $(z(\mathbf{D}),\mathbf{x}(\mathbf{D}),\mathbf{w}(\mathbf{D}))$ are the optimal recourse variables for demand realization $\mathbf{D}$. Let us define the following variables

$$\begin{pmatrix} z\\ \mathbf{x}\\ \mathbf{w}\end{pmatrix}=\mathbb{E}_f\left(\begin{pmatrix} z(\tilde{\mathbf{D}})\\ \mathbf{x}(\tilde{\mathbf{D}})\\ \mathbf{w}(\tilde{\mathbf{D}})\end{pmatrix}\right)$$

$$\begin{pmatrix}\mathbf{y}_{zs}^\top & \mathbf{y}_{zo}^\top\\ \mathbf{Y}_{xs} & \mathbf{Y}_{xo}\\ \mathbf{Y}_{ws} & \mathbf{Y}_{wo}\end{pmatrix}=\mathbb{E}_f\left(\begin{pmatrix} z(\tilde{\mathbf{D}})\\ \mathbf{x}(\tilde{\mathbf{D}})\\ \mathbf{w}(\tilde{\mathbf{D}})\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{D}}_s\\ \tilde{\mathbf{D}}_o\end{pmatrix}^\top\right)$$

$$\bar{\mathbf{X}}=\mathbb{E}_f\left(\mathbf{x}(\mathbf{D})\mathbf{x}(\mathbf{D})^\top\right)$$

$$\hat{\mathbf{X}}=\mathbb{E}_f\left(z(\mathbf{D})\mathbf{x}(\mathbf{D})\mathbf{x}(\mathbf{D})^\top\right).$$

Also define the constants

$$\boldsymbol{\Sigma}+\mathbf{m}\mathbf{m}^\top=\begin{pmatrix}\mathbf{Q}_{ss} & \mathbf{Q}_{so}^\top\\ \mathbf{Q}_{so} & \mathbf{Q}_{oo}\end{pmatrix}.$$

Note that

$$\begin{pmatrix}1\\ \mathbf{D}_s\\ \mathbf{D}_o\\ z(\mathbf{D})\\ \mathbf{x}(\mathbf{D})\\ \mathbf{w}(\mathbf{D})\end{pmatrix}\begin{pmatrix}1\\ \mathbf{D}_s\\ \mathbf{D}_o\\ z(\mathbf{D})\\ \mathbf{x}(\mathbf{D})\\ \mathbf{w}(\mathbf{D})\end{pmatrix}^\top=\begin{pmatrix}1 & \mathbf{D}_s^\top & \mathbf{D}_o^\top & z(\mathbf{D}) & \mathbf{x}(\mathbf{D})^\top & \mathbf{w}(\mathbf{D})^\top\\ \mathbf{D}_s & \mathbf{D}_s\mathbf{D}_s^\top & \mathbf{D}_s\mathbf{D}_o^\top & \mathbf{D}_s z(\mathbf{D}) & \mathbf{D}_s\mathbf{x}(\mathbf{D})^\top & \mathbf{D}_s\mathbf{w}(\mathbf{D})^\top\\ \mathbf{D}_o & \mathbf{D}_o\mathbf{D}_s^\top & \mathbf{D}_o\mathbf{D}_o^\top & \mathbf{D}_o z(\mathbf{D}) & \mathbf{D}_o\mathbf{x}(\mathbf{D})^\top & \mathbf{D}_o\mathbf{w}(\mathbf{D})^\top\\ z(\mathbf{D}) & z(\mathbf{D})\mathbf{D}_s^\top & z(\mathbf{D})\mathbf{D}_o^\top & z(\mathbf{D}) & \mathbf{w}(\mathbf{D})^\top & \mathbf{w}(\mathbf{D})^\top\\ \mathbf{x}(\mathbf{D}) & \mathbf{x}(\mathbf{D})\mathbf{D}_s^\top & \mathbf{x}(\mathbf{D})\mathbf{D}_o^\top & \mathbf{w}(\mathbf{D}) & \mathbf{x}(\mathbf{D})\mathbf{x}(\mathbf{D})^\top & z(\mathbf{D})\mathbf{x}(\mathbf{D})\mathbf{x}(\mathbf{D})^\top\\ \mathbf{w}(\mathbf{D}) & \mathbf{w}(\mathbf{D})\mathbf{D}_s^\top & \mathbf{w}(\mathbf{D})\mathbf{D}_o^\top & \mathbf{w}(\mathbf{D}) & z(\mathbf{D})\mathbf{x}(\mathbf{D})\mathbf{x}(\mathbf{D})^\top & z(\mathbf{D})\mathbf{x}(\mathbf{D})\mathbf{x}(\mathbf{D})^\top\end{pmatrix}$$

where we use the fact that $z(\mathbf{D})^2 = z(\mathbf{D})$, $z(\mathbf{D})\mathbf{x}(\mathbf{D}) = \mathbf{w}(\mathbf{D})$, $\mathbf{z}(\mathbf{D})\mathbf{w}(\mathbf{D}) = z(\mathbf{D})^2\mathbf{x}(\mathbf{D}) = z(\mathbf{D})\mathbf{x}(\mathbf{D}) = \mathbf{w}(\mathbf{D})$, $\mathbf{x}(\mathbf{D})\mathbf{w}(\mathbf{D})^\top = z(\mathbf{D})\mathbf{x}(\mathbf{D})\mathbf{x}(\mathbf{D})^\top$, $\mathbf{w}(\mathbf{D})\mathbf{w}(\mathbf{D})^\top = z(\mathbf{D})^2\mathbf{x}(\mathbf{D})\mathbf{x}(\mathbf{D})^\top = z(\mathbf{D})\mathbf{x}(\mathbf{D})\mathbf{x}(\mathbf{D})^\top$. Taking the expectation on both sides, we have that

$$\begin{pmatrix} 1 & \mathbf{m}_s^\top & \mathbf{m}_o^\top & z & \mathbf{x}^\top & \mathbf{w}^\top \\ \mathbf{m}_s & \mathbf{Q}_{ss} & \mathbf{Q}_{so}^\top & \mathbf{y}_{zs} & \mathbf{Y}_{xs}^\top & \mathbf{Y}_{ws}^\top \\ \mathbf{m}_o & \mathbf{Q}_{so} & \mathbf{Q}_{oo} & \mathbf{y}_{zo} & \mathbf{Y}_{xo}^\top & \mathbf{Y}_{wo}^\top \\ z & \mathbf{y}_{zs}^\top & \mathbf{y}_{zo}^\top & z & \mathbf{w}^\top & \mathbf{w}^\top \\ \mathbf{x} & \mathbf{Y}_{xs} & \mathbf{Y}_{xo} & \mathbf{w} & \bar{\mathbf{X}} & \hat{\mathbf{X}} \\ \mathbf{w} & \mathbf{Y}_{ws} & \mathbf{Y}_{wo} & \mathbf{w} & \hat{\mathbf{X}} & \hat{\mathbf{X}} \end{pmatrix} \succeq 0,$$

and that

$$\begin{pmatrix} 1 & z & \mathbf{x}^\top & \mathbf{w}^\top \\ z & z & \mathbf{w}^\top & \mathbf{w}^\top \\ \mathbf{x} & \mathbf{w} & \bar{\mathbf{X}} & \hat{\mathbf{X}} \\ \mathbf{w} & \mathbf{w} & \hat{\mathbf{X}} & \hat{\mathbf{X}} \end{pmatrix} \in \mathrm{BQP}.$$

Note that the a linear relaxation of the BQP constraints is the following:

$$\begin{aligned} \mathbf{w} &\le \mathbf{x}, \\ \mathbf{w} &\le z \cdot \mathbf{e}, \\ -\mathbf{w} + \mathbf{x} + z \cdot \mathbf{e} &\le 1, \\ \bar{X}_{ii} &= x_i, \\ \bar{X}_{ij} &\le x_i, \\ -\bar{X}_{ij} + x_i + x_j &\le 1, \\ \hat{X}_{ii} &= w_i, \\ \hat{X}_{ij} &\le w_i, \\ -\hat{X}_{ij} + w_i + w_j &\le 1, \\ \hat{X}_{ij} &\le x_i, \\ -\hat{X}_{ij} + x_i + w_j &\le 1. \end{aligned}$$

Removing redundant constraints and taking the dual of this SDP gives the Lemma. ■