E-Companion

EC.1. A general cutting plane algorithm (GCP)

Algorithm 2 GCP algorithm

1. Initialize: $\hat{\mathcal{O}} \leftarrow \widetilde{\Xi}$ contains the potential $\epsilon-$optimal solutions of $\boldsymbol{q}$;
2. $\tilde{\boldsymbol{q}} \leftarrow \boldsymbol{q}(\boldsymbol{w}, \hat{\mathbb{P}}_T)$ is the updated solution for the buyer's order;
3. Set upper bound: $\overline{v} \leftarrow (\boldsymbol{w}-\boldsymbol{c})^\top \tilde{\boldsymbol{q}}$. Set lower bound: $\underline{v} \leftarrow \min\left\{(\boldsymbol{w}-\boldsymbol{c})^\top \boldsymbol{q} : \boldsymbol{q} \in \hat{\mathcal{O}}\right\}$;
4. while $\overline{v} - \underline{v} > \varepsilon$ do

   Take some $\boldsymbol{q} \in \hat{\mathcal{O}}$ and check feasibility of Eq.(26) under $\boldsymbol{q}$;

   if Eq.(26) is infeasible, then update $\hat{\mathcal{O}} \leftarrow \hat{\mathcal{O}} - \chi_{\leq}(\boldsymbol{q})$;

   else, update $\hat{\mathcal{O}} \leftarrow \hat{\mathcal{O}} - \chi(\boldsymbol{q})$; if $(\boldsymbol{w}-\boldsymbol{c})^\top \boldsymbol{q} < \overline{v}$, then $\overline{v} = (\boldsymbol{w}-\boldsymbol{c})^\top \boldsymbol{q}$, and $\tilde{\boldsymbol{q}} \leftarrow \boldsymbol{q}$.

   Update $\underline{v} \leftarrow \min\left\{(\boldsymbol{w}-\boldsymbol{c})^\top \boldsymbol{q} : \boldsymbol{q} \in \hat{\mathcal{O}}\right\}$
5. Output: $\tilde{\boldsymbol{q}}$ as the optimal solution of buyer's order.

EC.2. Proofs

EC.2.1. Proof of Theorem 1

Theorem 1 is a special case of Theorem 4 when $n=1$. Please see the proof of Theorem 4.

EC.2.2. Proof of Theorem 2

Let $F_e(x)$, $F_T(x)$ and $F(x)$ denote the cumulative distribution functions of $\hat{\mathbb{P}}_T^e$, $\hat{\mathbb{P}}_T$ and $\overline{\mathbb{P}}$, respectively. When $n=1$, the type-2 Wasserstein distance has a special form

$$\begin{aligned} W_2^2(\hat{\mathbb{P}}_T^e, \overline{\mathbb{P}}) &= \int_0^1 \left(F_e^{-1}(\rho) - F^{-1}(\rho)\right)^2 d\rho \\ &= \int_0^1 \left(F_e^{-1}(\rho) - F_T^{-1}(\rho) + F_T^{-1}(\rho) - F^{-1}(\rho)\right)^2 d\rho. \end{aligned}$$

Next, we will show that for any $\rho \in [0,1]$, $\left(F_e^{-1}(\rho) - F_T^{-1}(\rho)\right) \cdot \left(F_T^{-1}(\rho) - F^{-1}(\rho)\right) \geq 0$. That is,

Case 1: when $F_e^{-1}(\rho) \leq F_T^{-1}(\rho)$, we have $F_T^{-1}(\rho) \leq F^{-1}(\rho)$;

Case 2: when $F_e^{-1}(\rho) \geq F_T^{-1}(\rho)$, we have $F_T^{-1}(\rho) \geq F^{-1}(\rho)$.

First, let us consider Case 1. Assume that for some $\rho \in [0,1]$, we have $F_e^{-1}(\rho) \leq F_T^{-1}(\rho)$ and $F_T^{-1}(\rho) > F^{-1}(\rho)$. We can find a new distribution $\hat{\mathbb{P}}_T'$ that satisfies:

$$F_T'^{-1}(\rho) = \begin{cases} F_T^{-1}(\rho), & if \quad F_e^{-1}(\rho) \leq F_T^{-1}(\rho), F_T^{-1}(\rho) \leq F^{-1}(\rho) \\ \max\left\{F_e^{-1}(\rho), F^{-1}(\rho)\right\}, & if \quad F_e^{-1}(\rho) \leq F_T^{-1}(\rho), F_T^{-1}(\rho) > F^{-1}(\rho) \end{cases}.$$

Note that the distribution $\hat{\mathbb{P}}_T'$ is a feasible center because it is compatible with the order conditions $\mathbb{P} \in \mathcal{D}_T^o$. Since for any $\rho \in [0,1]$, $0 \leq F_T'^{-1}(\rho) - F_e^{-1}(\rho) \leq F_T^{-1}(\rho) - F_e^{-1}(\rho)$, we have $W_2^2\left(\hat{\mathbb{P}}_T^e, \hat{\mathbb{P}}_T'\right) <$

$W_2^2\left(\hat{\mathbb{P}}_T^e, \hat{\mathbb{P}}_T\right)$. Therefore, $\hat{\mathbb{P}}_T$ is not the optimal solution of Problem (5). So the assumption is false. And we have if $F_e^{-1}(\rho) \leq F_T^{-1}(\rho)$, then $F_T^{-1}(\rho) \leq F^{-1}(\rho)$. Similarly, for Case 2, we have

$$W_2^2\left(\hat{\mathbb{P}}_T^e, \overline{\mathbb{P}}\right) = \int_0^1 \left(F_e^{-1}(\rho) - F_T^{-1}(\rho) + F_T^{-1}(\rho) - F^{-1}(\rho)\right)^2 d\rho \tag{EC.1}$$

$$\geq \int_0^1 \left(F_e^{-1}(\rho) - F_T^{-1}(\rho)\right)^2 d\rho + \int_0^1 \left(F_T^{-1}(\rho) - F^{-1}(\rho)\right)^2 d\rho \tag{EC.2}$$

$$\geq W_2^2\left(\hat{\mathbb{P}}_T, \overline{\mathbb{P}}\right). \tag{EC.3}$$

EC.2.3. Proof of Theorem 3

When joint distributions $\mathbb{P}_1, \mathbb{P}_2$ have independent marginal distributions $\{\mathbb{P}_1^i\}_{i\in[n]}, \{\mathbb{P}_2^i\}_{i\in[n]}$ respectively, then $W_2^2(\mathbb{P}_1, \mathbb{P}_2) = \sum_{i\in[n]} W_2^2(\mathbb{P}_1^i, \mathbb{P}_2^i)$ (Nguyen et al. 2023). That is, the Wasserstein distance between two probability distributions with independent marginals equal to the sum of the Wasserstein distances between their marginals. Note that $\times_{i\in[n]}\hat{\mathbb{P}}_T^i$ is a feasible solution to Problem (IM-center), so we have $W_2^2\left(\hat{\mathbb{Q}}_T, \hat{\mathbb{Q}}_T^e\right) \leq W_2^2\left(\times_{i\in[n]}\hat{\mathbb{P}}_T^i, \hat{\mathbb{Q}}_T^e\right) = \sum_{i\in[n]} W_2^2\left(\hat{\mathbb{P}}_T^i, \hat{\mathbb{P}}_T^{e,i}\right)$; As the marginal $\hat{\mathbb{Q}}_T^i$ is feasible to Problem (i-center) for all $i \in [n]$, and $W_2^2\left(\hat{\mathbb{Q}}_T, \hat{\mathbb{Q}}_T^e\right) = \sum_{i\in[n]} W_2^2\left(\hat{\mathbb{Q}}_T^i, \hat{\mathbb{P}}_T^{e,i}\right) \geq \sum_{i\in[n]} W_2^2\left(\hat{\mathbb{P}}_T^i, \hat{\mathbb{P}}_T^{e,i}\right) = W_2^2\left(\times_{i\in[n]}\hat{\mathbb{P}}_T^i, \hat{\mathbb{Q}}_T^e\right)$. Thus, $\times_{i\in[n]}\hat{\mathbb{P}}_T^i$ is optimal to problem (IM-center).

EC.2.4. Proof of Theorem 4

Based on Lagrangian duality, we reformulate Problem (6) into:

$$\sup_{\boldsymbol{\mu},\boldsymbol{\gamma}} \left\{ \frac{1}{T}\sum_{t=1}^{T} \gamma_t : \| \boldsymbol{\xi} - \boldsymbol{\xi}^t \|_2^2 + \sum_{i\in[n]}\sum_{t'=1}^{T} \mu_i^{t'} \left(\mathbf{1}\left(\xi_i \leq q_i^{t'}\right) - \frac{s_i - w_i^{t'}}{s_i} \right) \geq \gamma_t, \forall \boldsymbol{\xi} \in \Xi, t \in [T] \right\}. \tag{EC.4}$$

We first establish strong duality for Problem (11) and Problem (EC.4). Take Problem (EC.4) to be the primal (P) and Problem (6) is the dual (D). We let $\widetilde{\boldsymbol{c}} = \left(\left(-\frac{s_i - w_i^{t'}}{s_i}\right)_{i\in[n],t'\in[T]}, 1/T, \ldots, 1/T\right)$, $\boldsymbol{x} = \left(\left(\mu_i^{t'}\right)_{i\in[n],t'\in[T]}, (\gamma_t)_{t\in[T]}\right)$, and $a_t(\boldsymbol{\xi}) = \left(\left(\mathbf{1}\left(\xi_i \leq q_i^{t'}\right)\right)_{i\in[n],t'\in[T]}, \mathbf{0}_{t-1}, -1, \mathbf{0}_{T-t}\right)$, where $\mathbf{0}_t$ denotes a vector with t 0's. And let $b_t(\boldsymbol{\xi}) = -\| \boldsymbol{\xi} - \boldsymbol{\xi}^t \|_2^2$ for all $\boldsymbol{\xi} \in \Xi$ and $t \in [T]$. Then, Problem (P) can be rewritten as $\sup_{\boldsymbol{x}} \left\{ \widetilde{\boldsymbol{c}}^\top \boldsymbol{x} : a_t(\boldsymbol{\xi})^\top \boldsymbol{x} \geq b_t(\boldsymbol{\xi}), \forall \boldsymbol{\xi} \in \Xi, t \in [T] \right\}$. We define the following cone:

$$S = \operatorname{cone}\left\{ \left(a_t(\boldsymbol{\xi}), b_t(\boldsymbol{\xi})\right) : \boldsymbol{\xi} \in \Xi, t \in [T] \right\} \subset \mathbb{R}^{nT+T+1}.$$

According to HETTICH and Kortanek (1993), suppose val(P) is finite and S is closed, we have val(P)=val(D). First, because (P) is always feasible, we have val(P)$> -\infty$. And because (D) is always feasible ($\mathcal{D}_T^o$ is nonempty since $\overline{\mathbb{P}} \in \mathcal{D}_T^o$), we have val(D)$< \infty$. By weak duality, we obtain val(P)$\leq$val(D). Therefore, val(P) is finite. Second, notice that the set $\{a_t(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi, t \in [T]\}$ is finite. Since Ξ is compact and $b_t(\cdot)$ is continuous, the set $\{b_t(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi, t \in [T]\}$ is also compact since the continuous image of a compact set is compact. Thus, S is closed and the strong duality holds.

Based on the set partition and the duality theory of linear programming, we obtain the model

$$\text{(DLP)} \inf_{\boldsymbol{\beta}} \sum_{t=1}^{T}\sum_{m=1}^{M} \beta_{mt} \parallel \boldsymbol{\xi}_{mt}^{*} - \boldsymbol{\xi}^{t} \parallel_2^2 \tag{EC.5a}$$

$$\text{s.t.} \sum_{m=1}^{M} \beta_{mt} = 1/T, \forall t \in [T], \tag{EC.5b}$$

$$\sum_{t=1}^{T}\sum_{m=1}^{M} \beta_{mt} z_{im}\left(\boldsymbol{q}^{t'}\right) = \frac{s_i - w_i^{t'}}{s_i}, \forall i \in [n], t' \in [T], \tag{EC.5c}$$

$$\beta_{mt} \geq 0, \forall t \in [T], m \in [M]. \tag{EC.5d}$$

We now prove that $\hat{\mathbb{P}}_T$ defined by Eq.(22): $\hat{\mathbb{P}}_T = \sum_{t=1}^{T}\sum_{m=1}^{M} \beta_{mt}^{*}\delta_{\boldsymbol{\xi}_{mt}^{*}}$ is feasible and optimal for Problem (6). By the definition of $\boldsymbol{\xi}_{mt}^{*}$, we have $\boldsymbol{\xi}_{mt}^{*} \in \Xi$. By Eq. (EC.5a), $\hat{\mathbb{P}}_T = \sum_{t=1}^{T}\sum_{m=1}^{M} \beta_{mt}^{*}\delta_{\boldsymbol{\xi}_{mt}^{*}}$ is a valid probability distribution supported on Ξ. Furthermore, by Eq. (EC.5c), $\hat{\mathbb{P}}_T$ satisfies the buyer's first-order conditions. Thus, we see that $\hat{\mathbb{P}}_T \in \mathcal{D}_T^o$ and so $\hat{\mathbb{P}}_T$ is feasible for Problem (6). At the same time, according to the strong duality, the optimal values of Problem (6) and Problem (DLP) are equal. Thus, we have $\inf_{\mathbb{P}\in\mathcal{D}_T^o} W_2^2\left(\mathbb{P}, \hat{\mathbb{P}}_T^e\right) = \sum_{t=1}^{T}\sum_{m=1}^{M} \beta_{mt}^{*} \parallel \boldsymbol{\xi}_{mt}^{*} - \boldsymbol{\xi}^{t} \parallel_2^2$. So, $\hat{\mathbb{P}}_T$ achieves the minimum in Problem (6) and so is an optimal solution to Problem (6).

EC.2.5. Proof of Proposition 1

By the triangle inequality, we have $W_2\left(\overline{\mathbb{P}}, \hat{\mathbb{P}}_T\right) \leq W_2\left(\overline{\mathbb{P}}, \hat{\mathbb{P}}_T^e\right) + W_2\left(\hat{\mathbb{P}}_T^e, \hat{\mathbb{P}}_T\right)$. Since $\hat{\mathbb{P}}_T$ satisfies $W_2\left(\hat{\mathbb{P}}_T^e, \hat{\mathbb{P}}_T\right) \leq W_2\left(\overline{\mathbb{P}}, \hat{\mathbb{P}}_T^e\right)$ by construction, we have

$$W_2\left(\overline{\mathbb{P}}, \hat{\mathbb{P}}_T\right) \leq 2W_2\left(\overline{\mathbb{P}}, \hat{\mathbb{P}}_T^e\right). \tag{EC.6}$$

Fournier and Guillin (2015) give a concentration inequality for $\overline{\mathbb{P}}^T\left\{W_2\left(\overline{\mathbb{P}}, \hat{\mathbb{P}}_T^e\right) \geq \theta\right\}$. Combining with Eq.(EC.6), we have:

$$\overline{\mathbb{P}}^T\left\{W_2\left(\overline{\mathbb{P}}, \hat{\mathbb{P}}_T\right) \geq \theta\right\} \leq \begin{cases} c_1 \exp\left(-c_2 T\left(\frac{\theta}{2}\right)^{\max\{n,4\}}\right), & \frac{\theta}{2} \leq 1, \\ c_1 \exp\left(-c_2 T\left(\frac{\theta}{2}\right)^{a}\right), & \frac{\theta}{2} > 1, \end{cases} \tag{EC.7}$$

for all $T \geq 1$, $d \neq 4$,and $\theta > 0$, where c_1, c_2 are positive constants that only depend on n, a. Eq.(EC.7) provides a priori estimate of the probability that the unknown true distribution $\overline{\mathbb{P}}$ resides outside of the ambiguity set $\mathcal{D}_\theta\left(\hat{\mathbb{P}}_T\right)$. Based on this, let $\{\alpha_T\}_{T\geq 1}$ be a sequence of confidence levels $\alpha_T \in (0,1)$ such that $\sum_{T=1}^{\infty} \alpha_T < \infty$. For each $T \geq 1$, we have the confidence parameter:

$$\theta_T = \begin{cases} 2\left(\frac{\log\left(c_1\alpha_T^{-1}\right)}{c_2 T}\right)^{1/\max\{n,4\}}, & T \geq \frac{\log\left(c_1\alpha_T^{-1}\right)}{c_2}, \\ 2\left(\frac{\log\left(c_1\alpha_T^{-1}\right)}{c_2 T}\right)^{1/a}, & T < \frac{\log\left(c_1\alpha_T^{-1}\right)}{c_2}. \end{cases} \tag{EC.8}$$

According to Eq.(EC.7), which ensures via the definition of θ_T in (EC.8) that $\overline{\mathbb{P}}^T\left\{\overline{\mathbb{P}} \in \mathcal{D}_{\theta_T}\left(\hat{\mathbb{P}}_T\right)\right\} \geq 1 - \alpha_T$, we have $v\left(\boldsymbol{w}^*(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)), \overline{\mathbb{P}}\right) \geq v^*\left(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)\right)$ with probability $1 - \alpha_T$.

EC.2.6. Proof of Proposition 2

To prove the asymptotic consistency, we first give the following lemma.

Lemma EC.1. (Convergence of distributions) Suppose that Assumptions 1 and 2 hold and that $\alpha_T \in (0,1)$ satisfies $\sum_{T=1}^{\infty} \alpha_T < \infty$, and $\lim_{T\to\infty} \theta_T = 0$, then any sequence $\mathbb{P}_T \in \mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)$ converges under the Wasserstein metric to $\overline{\mathbb{P}}$ almost surely with respect to $\overline{\mathbb{P}}^{\infty}$, that is

$$\overline{\mathbb{P}}^{\infty}\left\{\lim_{T\to\infty} W_2\left(\overline{\mathbb{P}},\mathbb{P}_T\right)=0\right\}=1.$$

Proof. As $\mathbb{P}_T \in \mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)$, the triangle inequality for the Wasserstein metric ensures that $W_2\left(\overline{\mathbb{P}},\mathbb{P}_T\right) \le W_2\big(\overline{\mathbb{P}},\hat{\mathbb{P}}_T\big) + W_2\big(\mathbb{P}_T,\hat{\mathbb{P}}_T\big) \le W_2\big(\overline{\mathbb{P}},\hat{\mathbb{P}}_T\big) + \theta_T$. Proposition (1) implies that $\overline{\mathbb{P}}^T\left\{W_2\big(\overline{\mathbb{P}},\hat{\mathbb{P}}_T\big)\le\theta_T\right\}\ge 1-\alpha_T$, thus we have $\overline{\mathbb{P}}^T\left\{W_2\left(\overline{\mathbb{P}},\mathbb{P}_T\right)\le 2\theta_T\right\}\ge 1-\alpha_T$. As $\sum_{T=1}^{\infty}\alpha_T<\infty$, the Borel–Cantelli Lemma further implies that $\overline{\mathbb{P}}^{\infty}\left\{W_2\left(\overline{\mathbb{P}},\mathbb{P}_T\right)\le 2\theta_T \text{ for all sufficiently large } T\right\}=1$. Finally, as $\lim_{T\to\infty}\theta_T=0$, we conclude that $\lim_{T\to\infty} W_2\left(\overline{\mathbb{P}},\mathbb{P}_T\right)=0$ almost surely.

Next, we define the value functions $v\left(\boldsymbol{w},\boldsymbol{q},\mathbb{P}\right)=\left\{\left(\boldsymbol{w}-\boldsymbol{c}\right)^{\top}\boldsymbol{q}:\mathbb{P}^i\left(\xi_i\le q_i\right)\ge 1-w_i/s_i,\forall i\in[n]\right\}$, and $v\left(\boldsymbol{w},\mathbb{P}\right)=\min_{\boldsymbol{q}\ge 0} v\left(\boldsymbol{w},\boldsymbol{q},\mathbb{P}\right)$. Let $\boldsymbol{\lambda}=\left(\lambda_i\right)_{i\in[n]}$and $h\left(\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{q},\boldsymbol{\xi}\right)=\sum_{i\in[n]}\left(w_i-c_i\right)q_i - \sum_{i\in[n]}\lambda_i\left(\mathbf{1}\left(\xi_i\le q_i\right)-1+w_i/s_i\right)$. We have $v\left(\boldsymbol{w},\mathbb{P}\right)=\inf_{\boldsymbol{q}\ge 0}\sup_{\boldsymbol{\lambda}\ge 0}\mathbb{E}_{\mathbb{P}}\left[h\left(\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{q},\boldsymbol{\xi}\right)\right]$. According to Mohajerin Esfahani and Kuhn (2018), as $h\left(\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{q},\boldsymbol{\xi}\right)$ is upper semi-continuous in $\boldsymbol{\xi}$ and $\mid h\left(\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{q},\boldsymbol{\xi}\right)\mid\le L\left(1+\parallel\boldsymbol{\xi}\parallel\right)$ for some $L>0$, there exists a non-increasing sequence of Lipschitz continuous functions $h_k\left(\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{q},\boldsymbol{\xi}\right),k\in\mathbb{N}$. such that $h\left(\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{q},\boldsymbol{\xi}\right)=\lim_{k\to\infty}h_k\left(\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{q},\boldsymbol{\xi}\right)$. Let $v_k\left(\boldsymbol{w},\mathbb{P}\right)=\inf_{\boldsymbol{q}\ge 0}\sup_{\boldsymbol{\lambda}\ge 0}\mathbb{E}_{\mathbb{P}}\left[h_k\left(\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{q},\boldsymbol{\xi}\right)\right]$, then, we have $v\left(\boldsymbol{w},\mathbb{P}\right)=\lim_{k\to\infty}v_k\left(\boldsymbol{w},\mathbb{P}\right)$.

Proof of Part (a). As $\boldsymbol{w}^*(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T))\in[\boldsymbol{c},\boldsymbol{s}]$, we have $v\big(\boldsymbol{w}^*(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)),\overline{\mathbb{P}}\big)\le v^*\left(\overline{\mathbb{P}}\right)$. Moreover, Proposition 1 implies that $\overline{\mathbb{P}}^T\left\{v^*\big(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)\big)\le v\big(\boldsymbol{w}^*(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)),\overline{\mathbb{P}}\big)\le v^*\left(\overline{\mathbb{P}}\right)\right\}\ge \overline{\mathbb{P}}^T\left\{\overline{\mathbb{P}}\in\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)\right\}\ge 1-\alpha_T$. As $\sum_{T=1}^{\infty}\alpha_T<\infty$, according to the Borel-Cantelli Lemma we have $\overline{\mathbb{P}}^{\infty}\left\{v^*\big(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)\big)\le v\big(\boldsymbol{w}^*(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)),\overline{\mathbb{P}}\big)\le v^*\left(\overline{\mathbb{P}}\right) \text{ for all sufficiently large } T\right\}=1$.

To show that $v^*\big(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)\big)\to v^*\left(\overline{\mathbb{P}}\right)$, it remains to show that $\liminf_{T\to\infty} v^*\big(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)\big)\ge v^*\left(\overline{\mathbb{P}}\right)$. Let $\boldsymbol{w}^*\left(\overline{\mathbb{P}}\right)$ be the optimal solution the Problem (23) under $\overline{\mathbb{P}}$. $\boldsymbol{w}^*\left(\overline{\mathbb{P}}\right)$ exists because of the continuity of $v\left(\boldsymbol{w},\mathbb{P}\right)$ and the compactness of the feasible region. For each $T\ge 1$, we can find a $\delta-$optimal distribution $\mathbb{P}_T^{\delta}\in\mathcal{D}_{\theta_T}\left(\hat{\mathbb{P}}_T\right)$ such that $v\left(\boldsymbol{w}^*\left(\overline{\mathbb{P}}\right),\mathbb{P}_T^{\delta}\right)\le\inf_{\mathbb{P}\in\mathcal{D}_{\theta_T}\left(\hat{\mathbb{P}}_T\right)}v\left(\boldsymbol{w}^*\left(\overline{\mathbb{P}}\right),\mathbb{P}\right)+\delta$. We then have

$$\begin{aligned}
\liminf_{T\to\infty} v^*\big(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)\big) &\ge \liminf_{T\to\infty}\inf_{\mathbb{P}\in\mathcal{D}_{\theta_T}\left(\hat{\mathbb{P}}_T\right)} v\left(\boldsymbol{w}^*\left(\overline{\mathbb{P}}\right),\mathbb{P}\right)\ge\liminf_{T\to\infty} v\left(\boldsymbol{w}^*\left(\overline{\mathbb{P}}\right),\mathbb{P}_T^{\delta}\right)-\delta\\
&=\liminf_{T\to\infty}\inf_{\boldsymbol{q}\ge 0}\sup_{\boldsymbol{\lambda}\ge 0}\mathbb{E}_{\mathbb{P}_T^{\delta}}\left[h\left(\boldsymbol{\lambda},\boldsymbol{w}^*\left(\overline{\mathbb{P}}\right),\boldsymbol{q},\boldsymbol{\xi}\right)\right]-\delta\\
&\ge\lim_{k\to\infty}\liminf_{T\to\infty}\inf_{\boldsymbol{q}\ge 0}\sup_{\boldsymbol{\lambda}\ge 0}\mathbb{E}_{\mathbb{P}_{\mathbb{T}}^{\delta}}\left[h_k\left(\boldsymbol{\lambda},\boldsymbol{w}^*\left(\overline{\mathbb{P}}\right),\boldsymbol{q},\boldsymbol{\xi}\right)\right]-\delta\\
&\ge\lim_{k\to\infty}\liminf_{T\to\infty}\inf_{\boldsymbol{q}\ge 0}\sup_{\boldsymbol{\lambda}\ge 0}\left(\mathbb{E}_{\overline{\mathbb{P}}}\left[h_k\left(\boldsymbol{\lambda},\boldsymbol{w}^*\left(\overline{\mathbb{P}}\right),\boldsymbol{q},\boldsymbol{\xi}\right)\right]-L_kW_2\left(\overline{\mathbb{P}},\mathbb{P}_{\mathbb{T}}^{\delta}\right)\right)-\delta\\
&=\lim_{k\to\infty}\inf_{\boldsymbol{q}\ge 0}\sup_{\boldsymbol{\lambda}\ge 0}\mathbb{E}_{\overline{\mathbb{P}}}\left[h_k\left(\boldsymbol{\lambda},\boldsymbol{w}^*\left(\overline{\mathbb{P}}\right),\boldsymbol{q},\boldsymbol{\xi}\right)\right]-\delta,\quad \overline{\mathbb{P}}^{\infty}\text{- almost surely}\\
&=\lim_{k\to\infty}v_k\left(\boldsymbol{w}^*\left(\overline{\mathbb{P}}\right),\overline{\mathbb{P}}\right)-\delta=v\left(\boldsymbol{w}^*\left(\overline{\mathbb{P}}\right),\overline{\mathbb{P}}\right)-\delta=v^*\left(\overline{\mathbb{P}}\right)-\delta.
\end{aligned}$$

In the above display, the third inequality holds because of the convergence of $h_k(\boldsymbol{\lambda}, \boldsymbol{w}, \boldsymbol{q}, \boldsymbol{\xi})$. The fourth inequality and second equality follow from Mohajerin Esfahani and Kuhn (2018). Since δ is arbitrary, we have $\liminf_{T\to\infty} v^*\big(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)\big) \geq v^*\left(\overline{\mathbb{P}}\right)$.

Proof of Part (b). Fix a realization of the stochastic process $\left\{\hat{\boldsymbol{\xi}}_N\right\}_{N\geq 1}$ such that $\lim_{T\to\infty} v^*\big(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)\big) = v^*\left(\overline{\mathbb{P}}\right)$ and $v^*\big(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)\big) \leq v\big(\boldsymbol{w}^*(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)), \overline{\mathbb{P}}\big) \leq v^*\left(\overline{\mathbb{P}}\right)$ for all sufficiently large T. From Part (a), these two conditions are satisfied $\overline{\mathbb{P}}^{\infty}$- almost surely. By the closeness of $[\boldsymbol{c}, \boldsymbol{s}]$, $\left\{\boldsymbol{w}^*(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T))\right\}_{T\geq 1}$ has an accumulation point $\overline{\boldsymbol{w}} \in [\boldsymbol{c}, \boldsymbol{s}]$. Thus, we have $v^*\left(\overline{\mathbb{P}}\right) \geq v\left(\overline{\boldsymbol{w}}, \overline{\mathbb{P}}\right) = \lim_{T\to\infty} v\big(\boldsymbol{w}^*(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)), \overline{\mathbb{P}}\big) \geq \lim_{T\to\infty} v^*\big(\mathcal{D}_{\theta_T}(\hat{\mathbb{P}}_T)\big) = v^*\left(\overline{\mathbb{P}}\right)$. We conclude that $v\left(\overline{\boldsymbol{w}}, \overline{\mathbb{P}}\right) = v^*\left(\overline{\mathbb{P}}\right)$.

EC.2.7. Proof of Proposition 3

Without loss of generality, assuming $\hat{\mathbb{P}}_T = \delta_{\mathbf{1}_n}$, $\widetilde{\Xi} = \{0,1\}^n$, and the order constraint is naturally satisfied. Then, for given $\boldsymbol{w}$, the seller's worst-case profit problem $\min_{\boldsymbol{q}\geq 0, \mathbb{P}\in\mathcal{D}_\theta(\hat{\mathbb{P}}_T)} \left(\boldsymbol{w}-\boldsymbol{c}\right)^\top \boldsymbol{q}\big(\boldsymbol{w}, \mathbb{P}\big)$, which can be reformulated as a single-level problem by the buyer's first-order conditions,

$$\min_{\boldsymbol{q}\geq 0} \left(\boldsymbol{w}-\boldsymbol{c}\right)^\top \boldsymbol{q} \qquad \text{s.t.} \int_{\boldsymbol{\xi}\in\Xi} \|\boldsymbol{\xi}-\mathbf{1}_n\|_2^2 \mathbb{P}\big(d\boldsymbol{\xi}\big) \leq \theta^2, \qquad \mathbb{P}^i\left(\xi_i \leq q_i\right) \geq 1 - w_i/s_i, \forall i \in [n].$$

Because $\widetilde{\Xi} = \{0,1\}^n$, q_i is either 0 or 1 for any $i \in [n]$. Following the same idea as LRDS, only vertices are potential mass transfer points. Thus, the above problem is equivalent to:

$$\min_{\boldsymbol{q}\geq 0} \left(\boldsymbol{w}-\boldsymbol{c}\right)^\top \boldsymbol{q} \qquad \text{s.t.} \sum_{i=1}^{n} (1-w_i/s_i)(q_i-1) \leq \theta^2.$$

This is essentially a classic backpack problem, which is known to be NP-hard.

EC.2.8. Proof of Proposition 4

Given w_i, we find the past wholesale prices $\overline{w_i}$, $\underline{w_i}$, the corresponding past buyer's orders are $\underline{q_{i,w}}$ and $\overline{q_{i,w}}$. According to the order conditions, we have $\mathbb{P}^i\big(\xi_i \leq \underline{q_{i,w}}\big) = 1 - \overline{w_i}/s_i \leq 1 - w_i/s_i$ holds for all $\mathbb{P} \in \mathcal{D}_T^o$, so we have $q_i \geq \underline{q_{i,w}}$. On the other hand, we also have $\mathbb{P}^i\left(\xi_i \leq \overline{q_{i,w}}\right) = 1 - \underline{w_i}/s_i > 1 - w_i/s_i$ holds for all $\mathbb{P} \in \mathcal{D}_T^o$, so we have $q_i < \overline{q_{i,w}}$. Thus, we have $q_i^* \subset \big[\underline{q_{i,w}}, \overline{q_{i,w}}\big)$.

Next, we will show that there is always a feasible $\mathbb{P}_{\boldsymbol{w},\boldsymbol{q}} \in \mathcal{D}_T^o$ for $\forall \boldsymbol{q} \in \times_{i=1}^n \big[\underline{q_{i,w}}, \overline{q_{i,w}}\big)$. To do this, for any $i \in [n]$, we establish the following marginal distribution $\widetilde{\mathbb{P}}^i_{\boldsymbol{w},\boldsymbol{q}}$ in Figure EC.1(a). It is easy to find that $\widetilde{\mathbb{P}}^i_{\boldsymbol{w},\boldsymbol{q}}$ satisfies the order constraints. Given w_i ($\overline{w_i} \geq w_i > \underline{w_i}$), for any $q_i \in \big[\underline{q_{i,w}}, \overline{q_{i,w}}\big)$, we can always find a marginal distribution $\mathbb{P}^i_{\boldsymbol{w},\boldsymbol{q}}$ in Figure EC.1(b) by transferring the masses $\left(\overline{w_i} - w_i\right)/s_i$ from $\overline{q_{i,w}}$ to q_i. Note that the distribution $\mathbb{P}^i_{\boldsymbol{w},\boldsymbol{q}}$ satisfies $q_i\left(w_i, \mathbb{P}^i_{\boldsymbol{w},\boldsymbol{q}}\right) = q_i$ and does not violate the order constraints. Therefore, for $\forall \boldsymbol{q} \in \times_{i=1}^n \big[\underline{q_{i,w}}, \overline{q_{i,w}}\big)$, we can always find a feasible distribution $\mathbb{P}_{\boldsymbol{w},\boldsymbol{q}} \in \mathcal{D}_T^o$, and the proposition is proved.

EC.2.9. Proof of Proposition 5

We can reformulate Eq.(27) in a more tractable form. To check the feasibility, we need to solve

$$\text{Feasi } (\boldsymbol{q}) \ : \theta\big(\boldsymbol{q}\big) = \min_{\mathbb{P}\in\mathcal{P}(\Xi)} W_2^2\big(\mathbb{P}, \hat{\mathbb{P}}_T\big) \text{ s.t. } \ \mathbb{P}^i\left(\xi_i \leq q_i\right) \geq 1 - w_i/s_i, \forall i \in [n]. \tag{EC.9}$$

Eq.(27) is feasible if and only if the optimal value $\theta\left(\boldsymbol{q}\right) \leq \theta^2$. Note that for any $\boldsymbol{q}' \in \chi_{\leq}\left(\boldsymbol{q}\right)$, Feasi $(\boldsymbol{q})$ is a relaxation of Feasi $(\boldsymbol{q}')$, so $\theta\left(\boldsymbol{q}'\right) \geq \theta\left(\boldsymbol{q}\right)$. Therefore, if $\boldsymbol{q}$ is infeasible, $\boldsymbol{q}'$ is also infeasible.

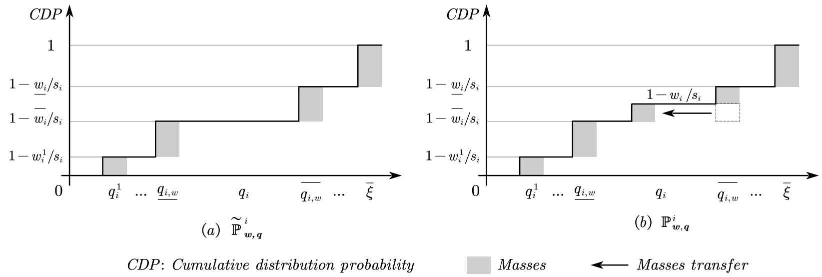

Figure EC.1 **The continuous feasibility for** $\boldsymbol{q}\in\times_{i=1}^n\left[\underline{q_{i,w}},\overline{q_{i,w}}\right)$

EC.2.10. Proof of Theorem 5

In Algorithm 1, we construct an $\epsilon-$net $\widetilde{\Xi}$ for $\epsilon=\frac{\sigma}{n\max_{i\in[n]}\{s_i-w_i\}}$ to find an $\sigma-$optimal solution of Problem (25). Given the optimal solution $\boldsymbol{q}\big(\boldsymbol{w},\mathcal{D}_\theta(\hat{\mathbb{P}}_T)\big)$, we can always find an order $\tilde{\boldsymbol{q}}^*$ in this $\epsilon-$net by Algorithm 1, and the seller's profit satisfies $(\boldsymbol{w}-\boldsymbol{c})^\top\mid\tilde{\boldsymbol{q}}^*-\boldsymbol{q}\big(\boldsymbol{w},\mathcal{D}_\theta(\hat{\mathbb{P}}_T)\big)\mid\leq\sigma$. The $\epsilon-$net only has a finite number of elements, and each iteration can determine the optimality for at least one element. Therefore, after a finite number of iterations, the algorithm converges.

EC.2.11. Proof of Proposition 6

Given w, the distribution $\mathbb{P}_{w,q}$ can be obtained by distorting the center distribution $\hat{\mathbb{P}}_T$.

i. For $q=\xi_w$, $\mathbb{P}_{w,q}$ can be the center distribution $\hat{\mathbb{P}}_T$, so $W_2^2\big(\mathbb{P}_{w,q},\hat{\mathbb{P}}_T\big)=0$. The Proposition holds.

ii. For any $\underline{q_w}\leq q<\xi_w$, the corresponding distribution $\mathbb{P}_{w,q}$ should satisfy $\mathbb{P}_{w,q}(\xi\leq q)=1-w/s$. Due to $\xi_w=\arg\min_{x\in\left(\hat{\xi}^k\right)_{k\in[K]}}\left\{x:\hat{\mathbb{P}}_T(\xi\leq x)\geq 1-w/s\right\}$, we have $\hat{\mathbb{P}}_T(\xi\leq q)<1-w/s$, so we need to transfer the masses from $\left\{\hat{\xi}^k:\hat{\xi}^k>q,k\in[K]\right\}$ to q. And we can achieve the minimum transportation cost in a greedy fashion (This idea is consistent with the continuous knapsack problem), that is, starting from the point closest to q on the right-hand side, we transfer the masses towards q until its cumulative distribution probability reaches $1-w/s$. Since ξ_w is the minimum location where the cumulative distribution probability reaches $1-w/s$, thus, all masses on the left side of ξ_w need to be fully transferred, those on the right side of ξ_w do not need to be transferred, and the masses at ξ_w needs to be partially transferred. The optimal transportation cost $W_2^2\big(\mathbb{P}_{w,q},\hat{\mathbb{P}}_T\big)$ is equal to $\sum_{k:q<\hat{\xi}^k<\xi_w}\eta^k\big(q-\hat{\xi}^k\big)^2+\big(1-w/s-\sum_{k:\hat{\xi}^k<\xi_w}\eta^k\big)(q-\xi_w)^2$, which is equal to $\sum_{k=1}^K\eta'_{\hat{\xi}^k}\left[\big(\hat{\xi}^k-q\big)^+\right]^2$.

EC.2.12. Proof of Theorem 6

From Proposition 4, q is continuously feasible for $\mathbb{P}\in\mathcal{D}_T^o$ within region $\left[\underline{q_w},\overline{q_w}\right)$. By Proposition 6, the minimum Wasserstein distance $W_2^2\big(\mathbb{P}_{w,q},\hat{\mathbb{P}}_T\big)$ is monotonically non-increasing concerning q. And when $q=\underline{q_w}$, the minimum Wasserstein distance $W_2^2\big(\mathbb{P}_{w,q},\hat{\mathbb{P}}_T\big)=\sum_{k=1}^K\eta'_{\hat{\xi}^k}\left[\big(\hat{\xi}^k-\underline{q_w}\big)^+\right]^2=\overline{\theta}(w)$.

i) when we set $\theta\geq\overline{\theta}(w)$, the buyer's worst-case order quantity q^* satisfies $q^*\leq\underline{q_w}$, and since the constraint $q\geq\underline{q_w}$, we have $q^*=\underline{q_w}$;

ii) when we set $\theta < \overline{\theta}(w)$, the buyer's worst-case order quantity q^* satisfies $q^* > \underline{q_w}$. To obtain the worst-case (the least) order quantity, q^* must be obtained at the maximum budget, that is $\sum_{k=1}^{K} \eta'_{\hat{\xi}^k} \left[\left(\hat{\xi}^k - q^* \right)^+ \right]^2 = \theta^2$.

EC.2.13. Proof of Proposition 7

Following the Proposition 4, we have $\underline{q_{i,w}} \le q_i^* < \overline{q_{i,w}}$ for all $i \in [n]$. Due to $\hat{\mathbb{Q}}_T \in \mathcal{D}_T^o \cap \mathcal{D}_\times$, it is easy to know that $\sum_{i=1}^{n} (w_i - c_i)\, q_i^* \le \sum_{i=1}^{n} (w_i - c_i)\, \xi_{i,w}$. Assume that $\boldsymbol{q}^* = (q_i^*)_{i\in[n]}$, and $\exists q_j^* : \xi_{j,w} < q_j^* < \overline{q_{j,w}}$ for some $j \in [n]$, we have $\mathbb{P}^j_{\boldsymbol{w},\boldsymbol{q}^*} \neq \hat{\mathbb{P}}^j_T$, so $W_2^2\left(\mathbb{P}^j_{\boldsymbol{w},\boldsymbol{q}^*}, \hat{\mathbb{P}}^j_T\right) > 0$. By Nguyen et al. (2023), the $W_2^2\left(\mathbb{P}, \hat{\mathbb{Q}}_T\right) = \sum_{i=1}^{n} W_2^2\left(\mathbb{P}^i, \hat{\mathbb{P}}^i_T\right)$ holds. Due to $W_2^2\left(\mathbb{P}_{\boldsymbol{w},\boldsymbol{q}^*}, \hat{\mathbb{Q}}_T\right) \le \theta^2$, we have $\sum_{i=1}^{j-1} W_2^2\left(\mathbb{P}^i_{\boldsymbol{w},\boldsymbol{q}^*}, \hat{\mathbb{P}}^i_T\right) + \sum_{i=j+1}^{n} W_2^2\left(\mathbb{P}^i_{\boldsymbol{w},\boldsymbol{q}^*}, \hat{\mathbb{P}}^i_T\right) + W_2^2\left(\mathbb{P}^j_{\boldsymbol{w},\boldsymbol{q}^*}, \hat{\mathbb{P}}^j_T\right) \le \theta^2$. Because $W_2^2\left(\hat{\mathbb{P}}^j_T, \hat{\mathbb{P}}^j_T\right) = 0 < W_2^2\left(\mathbb{P}^j_{\boldsymbol{w},\boldsymbol{q}^*}, \hat{\mathbb{P}}^j_T\right)$, so $q_i = q_i^*$ for all $i \in [n] \setminus \{j\}$ and $q_j = \xi_{j,w}$ are feasible for the model. But $\sum_{i=1}^{n} (w_i - c_i)\, q_i^* > \sum_{i=1}^{j-1} (w_i - c_i)\, q_i^* + \sum_{i=j+1}^{n} (w_i - c_i)\, q_i^* + (w_j - c_j)\, \xi_{j,w}$. Thus, the assumption is not valid, and we have $\underline{q_{i,w}} \le q_i^* \le \xi_{i,w}$ for all $i \in [n]$. Due to $W_2^2\left(\mathbb{P}_{\boldsymbol{w},\boldsymbol{q}}, \hat{\mathbb{Q}}_T\right) = \sum_{i\in[n]} W_2^2\left(\mathbb{P}^i_{\boldsymbol{w},\boldsymbol{q}}, \hat{\mathbb{P}}^i_T\right)$, so, we only need to formulate the minimum Wasserstein distance from the corresponding marginal distribution $\mathbb{P}^i_{\boldsymbol{w},\boldsymbol{q}}$ to the marginal center distribution $\hat{\mathbb{P}}^i_T$. Following the Proposition 6, we have $W_2^2\left(\mathbb{P}_{\boldsymbol{w},\boldsymbol{q}}, \hat{\mathbb{Q}}_T\right) = \sum_{i\in[n]} \sum_{k=1}^{K_i} \eta'_{\hat{\xi}_i^k} \left[\left(\hat{\xi}_i^k - q_i \right)^+ \right]^2$.

EC.3. LD Method for Solving the Augmented Center Distribution

We begin with the linear semi-infinite programming problem (LSP).

$$\sup_{\boldsymbol{\mu},\boldsymbol{\gamma}} \frac{1}{T} \sum_{t=1}^{T} \gamma_t - \sum_{i\in[n]} \sum_{t'=1}^{T} \mu_i^{t'} \frac{s_i - w_i^{t'}}{s_i} \text{s.t.} \parallel \boldsymbol{\xi} - \boldsymbol{\xi}^t \parallel_2^2 + \sum_{i\in[n]} \sum_{t'=1}^{T} \mu_i^{t'} \mathbf{1}\left(\xi_i \le q_i^{t'}\right) \ge \gamma_t, \forall \boldsymbol{\xi} \in \Xi, t \in [T]. \tag{EC.10a}$$

For each $i \in [n]$, we arrange the past wholesale prices in descending order $s_i \ge w_i^1 > ... > w_i^t > ... > w_i^T$. Then, the buyer's corresponding past order quantities in ascending order are $0 \le q_i^1 < ... < q_i^t < ... < q_i^T$. We set $q_i^0 = 0$ and $q_i^{T+1} = \overline{\xi}$ for all $i \in [n]$. The past orders $\{\boldsymbol{q}^t\}_{t\in[T]}$ partition Ξ into $(T+1)^n$ subsets $R_m = \left\{ \boldsymbol{\xi} : \xi_i \in \left(q_i^{m_i - 1}, q_i^{m_i} \right], \forall i \in [n] \right\}$ indexed by $m \in [M] = [(T+1)^n]$ where $(m_i)_{i=1}^n \in \{1, ..., T+1\}^n$. Then, for some $\boldsymbol{\xi} \in R_m$, $\mathbf{1}\left(\xi_i \le q_i^{t'}\right) = 0$ if $q_i^{m_i - 1} \ge q_i^{t'}$ and $\mathbf{1}\left(\xi_i \le q_i^{t'}\right) = 1$ if $q_i^{m_i} \le q_i^{t'}$. Thus, $\mathbf{1}\left(\xi_i \le q_i^{t'}\right)$ is constant for $\boldsymbol{\xi} \in R_m$. We define $z_{im}\left(\boldsymbol{q}^{t'}\right) = \mathbf{1}\left(\xi_i \le q_i^{t'} | \boldsymbol{\xi} \in R_m\right)$, and we represent R_m as $R_m = \{\boldsymbol{\xi} : A_m \boldsymbol{\xi} \le \boldsymbol{d}_m\}$ where A_m is a matrix and $\boldsymbol{d}_m$ is a vector of appropriate dimension. Problem (LSP) can then be reformulated as:

$$\sup_{\mu,\gamma} \left\{ \frac{1}{T} \sum_{t=1}^{T} \gamma_t : \parallel \boldsymbol{\xi} - \boldsymbol{\xi}^t \parallel_2^2 + \sum_{i\in[n]} \sum_{t'=1}^{T} \mu_i^{t'} \left(z_{im}\left(\boldsymbol{q}^{t'}\right) - \frac{s_i - w_i^{t'}}{s_i} \right) \ge \gamma_t, \forall \boldsymbol{\xi} \in R_m, t \in [T], m \in [M] \right\}, \tag{EC.11}$$

where the constraints are grouped by subsets of the partition. We will show that Problem (EC.11) is equivalent to the following convex optimization:

$$\inf \sum_{t=1}^{T} \sum_{m=1}^{M} \beta_{mt} \left(\boldsymbol{p}_{mt}/\beta_{mt} - \boldsymbol{\xi}^t \right)^\top \left(\boldsymbol{p}_{mt}/\beta_{mt} - \boldsymbol{\xi}^t \right) \tag{EC.12a}$$

$$\text{s.t.} \sum_{m=1}^{M} \beta_{mt} = 1/T, \forall t \in [T], \tag{EC.12b}$$

$$\sum_{t=1}^{T}\sum_{m=1}^{M}\beta_{mt}z_{im}\left(\boldsymbol{q}^{t'}\right)=\frac{s_i-w_i^{t'}}{s_i},\forall i\in[n],t'\in[T], \tag{EC.12c}$$

$$A_m\boldsymbol{p}_{mt}/\beta_{mt}\leq\boldsymbol{d}_s,\forall t\in[T],m\in[M], \tag{EC.12d}$$

$$\beta_{mt}\geq 0,\boldsymbol{p}_{mt}\in\mathbb{R}^n,\forall t\in[T],m\in[M]. \tag{EC.12e}$$

Next, we prove the equivalence between (EC.12) and (EC.11). Following Schur's complement, (EC.11) is equivalent to

$$\begin{array}{ll}\sup_{\mu,\gamma,\boldsymbol{\tau}} & \frac{1}{T}\sum_{t=1}^{T}\gamma_t\\ \text{s.t.} & \begin{bmatrix} I_{n\times n} & -\left(A_m^{\top}\boldsymbol{\tau}_{mt}+2\boldsymbol{\xi}^t\right)/2\\ -\left(A_m^{\top}\boldsymbol{\tau}_{mt}+2\boldsymbol{\xi}^t\right)/2 & \boldsymbol{\tau}_{mt}^{\top}\boldsymbol{d}_m+\boldsymbol{\xi}^{t\top}\boldsymbol{\xi}^t+\sum_{i\in[n]}\sum_{t'=1}^{T}\mu_i^{t'}\left(z_{im}\left(\boldsymbol{q}^{t'}\right)-\frac{s_i-w_i^{t'}}{s_i}\right)-\gamma_t\end{bmatrix}\succeq 0,\\ & \forall t\in[T],m\in[M],\\ & \boldsymbol{\tau}_{mt}\leq 0,\forall t\in[T],m\in[M],\end{array}$$

where $\boldsymbol{\tau}=\{\boldsymbol{\tau}_{mt}\}_{t\in[T],m\in[M]}$ and $I_{n\times n}$ stands for the identity matrix with size $n\times n$. The following Lagrangian dual to Problem (EC.11) then follows from SDP duality:

$$\inf\sum_{t=1}^{T}\sum_{m=1}^{M}I_{n\times n}\bullet Q_{mt}-2\boldsymbol{p}_{mt}^{\top}\boldsymbol{\xi}^t+\beta_{mt}\boldsymbol{\xi}^{t\top}\boldsymbol{\xi}^t \tag{EC.13a}$$

$$\text{s.t.}\sum_{m=1}^{M}\beta_{mt}=1/T,\forall t\in[T], \tag{EC.13b}$$

$$\sum_{t=1}^{T}\sum_{m=1}^{M}\beta_{mt}z_{im}\left(\boldsymbol{q}^{t'}\right)=\frac{s_i-w_i^{t'}}{s_i},\forall i\in[n],t'\in[T], \tag{EC.13c}$$

$$A_m\boldsymbol{p}_{mt}\leq\beta_{mt}\boldsymbol{d}_m,\forall t\in[T],m\in[M], \tag{EC.13d}$$

$$\begin{bmatrix} Q_{mt} & \boldsymbol{p}_{mt}\\ \boldsymbol{p}_{mt}^{\top} & \beta_{mt}\end{bmatrix}\succeq 0,\forall t\in[T],m\in[M], \tag{EC.13e}$$

$$Q_{mt}\in\mathbb{R}^{n\times n},\beta_{mt}\in\mathbb{R},\boldsymbol{p}_{mt}\in\mathbb{R}^n,\forall t\in[T],m\in[M], \tag{EC.13f}$$

where $\bullet$ is the Frobenius inner product, and $Q_{mt}\in\mathbb{R}^{n\times n},\beta_{mt}\in\mathbb{R},\boldsymbol{p}_{mt}\in\mathbb{R}^n$ are collectively the (matrix-valued) Lagrange multiplier to the constraints of Problem (EC.11). It is immediate that the relative interior of the feasible region of Problem (EC.11) is non-empty as $\boldsymbol{\gamma}$ is a free variable. Then, based on the weaker version of Shapiro (2001), Proposition 3.4, strong duality holds between Problems (EC.13) and (EC.11).

Let $(\cdot)^*$ denote the optimal value of $(\cdot)$. We will simplify the dual problem by showing that $Q_{mt}^*=\boldsymbol{p}_{mt}^*\boldsymbol{p}_{mt}^{*\top}/\beta_{mt}^*$ for all $t\in[T],m\in[M]$. In particular, we have $Q_{mt}^*=0$ when $\beta_{mt}^*=0$, since $\boldsymbol{p}_{mt}^*=0$ and $0/0=0$. Indeed, $\beta_{mt}^*=0$ implies $\boldsymbol{p}_{mt}^*=0$ and $Q_{mt}^*=0$ as $Q_{mt}\succeq 0$ must hold. When $\beta_{mt}^*>0$, Schur's complement shows that $Q_{mt}\succeq\boldsymbol{p}_{mt}\boldsymbol{p}_{mt}^{\top}/\beta_{mt}$ and so $Q_{mt}^*=\boldsymbol{p}_{mt}^*\boldsymbol{p}_{mt}^{*\top}/\beta_{mt}^*$ is an optimal solution. Substituting $Q_{mt}=\boldsymbol{p}_{mt}\boldsymbol{p}_{mt}^{\top}/\beta_{mt}$ in the objective of (EC.13), we obtain $\sum_{t=1}^{T}\sum_{m=1}^{M}\boldsymbol{p}_{mt}^{\top}\boldsymbol{p}_{mt}/\beta_{mt}-2\boldsymbol{p}_{mt}^{\top}\boldsymbol{\xi}^t+\beta_{mt}\boldsymbol{\xi}^{t\top}\boldsymbol{\xi}^t=\sum_{t=1}^{T}\sum_{m=1}^{M}\beta_{mt}\left(\boldsymbol{p}_{mt}/\beta_{mt}-\boldsymbol{\xi}^t\right)^{\top}\left(\boldsymbol{p}_{mt}/\beta_{mt}-\boldsymbol{\xi}^t\right)$, and so (EC.13) equals to (EC.12).

We now prove that $\hat{\mathbb{P}}_T$ defined by $\hat{\mathbb{P}}_T=\sum_{t=1}^{T}\sum_{m=1}^{M}\beta_{mt}^*\delta_{\boldsymbol{p}_{mt}^*/\beta_{mt}^*}$ is feasible and optimal for Problem (6). By Eq.(EC.12d), we have $\boldsymbol{p}_{mt}^*/\beta_{mt}^*\in\Xi$. By Eq.(EC.12b), $\hat{\mathbb{P}}_T=\sum_{t=1}^{T}\sum_{m=1}^{M}\beta_{mt}^*\delta_{\boldsymbol{p}_{mt}^*/\beta_{mt}^*}$

Table EC.1 Computational time (s) of LRDS for independent multi-product case

T \ n	10	20	30	50	80	100
10	0.25	0.4	0.7	1.1	1.6	2.1
20	0.7	1.5	2.3	3.6	5.5	7.5
30	1.5	3.2	4.5	7.7	11.2	15.4
40	2.5	5.2	7.3	13.1	19.8	27.9
50	4.3	8.9	12.4	21.3	31.6	42.0

Table EC.2 Computational time (s) of SOCP for solving the worst-case profit

(θ, T) \ n	10	20	30	50	80	100
(1, 10)	1.9	3.4	4.7	8.1	11.4	15.3
(1, 20)	2.0	3.6	4.9	8.4	12.6	17.4
(1, 30)	2.2	3.9	5.1	8.7	13.9	19.1
(2, 10)	1,9	3.2	4.8	7.9	11.8	14.9
(2, 20)	2.1	3.5	4.9	8.3	12.8	16.8
(2, 30)	2.3	3.9	5.2	8.9	13.6	18.2

is a valid probability distribution supported on Ξ. Furthermore, by Eq.(EC.12c), $\hat{\mathbb{P}}_T$ satisfies the retailer's first-order conditions since

$$\hat{\mathbb{P}}_T(\xi_i \le q_i^{t'}) = \sum_{m=1}^{M} \hat{\mathbb{P}}_T(\xi_i \le q_i^{t'} | \boldsymbol{\xi} \in R_m)\hat{\mathbb{P}}_T(\boldsymbol{\xi} \in R_m) = \sum_{t=1}^{T}\sum_{m=1}^{M} \beta_{mt}^* z_{im}(\boldsymbol{q}^{t'}) = \frac{s_i - w_i^{t'}}{s_i}.$$

Thus, we see that $\hat{\mathbb{P}}_T \in \mathcal{D}_T^o$ and so $\hat{\mathbb{P}}_T$ is feasible for (6). By the strong duality, the optimal values of (6) and (EC.12) are equal. We have $\inf_{\mathbb{P}\in\mathcal{D}_T^o} W_2^2(\mathbb{P}, \hat{\mathbb{P}}_T^e) = \sum_{t=1}^{T}\sum_{m=1}^{M} \beta_{mt}^* \left(\boldsymbol{p}_{mt}^*/\beta_{mt}^* - \boldsymbol{\xi}^t\right) \top \left(\boldsymbol{p}_{mt}^*/\beta_{mt}^* - \boldsymbol{\xi}^t\right)$. So, $\hat{\mathbb{P}}_T$ achieves the minimum of (6).

EC.4. Computational Performance of Large-scale Problems

For the independent multi-product case, we can obtain the center distribution for each product individually efficiently using the LRDS method, and establish the joint center distribution by Cartesian product operations. Then we solve the seller's worst-case profit using the SOCP reformulation, which can be efficiently solved by Gurobi for large instances, as shown in Tables EC.1 and EC.2.

For the dependent multi-product case, the computational efficiency of the PCP algorithm deteriorates as n increases. To address this issue, we propose approximating the dependent multi-product model with the independent case model. Tables EC.1 and Table EC.2 clearly exhibit superior computational efficiency. As shown in Table EC.3, ignoring correlations only has a minimal impact on the seller's worst-case profit, with a maximum degradation of 10% across all instances.

EC.5. Discussions on Model with Tolerance

In our problem, when introducing additional tolerance $\Delta \ge 0$ in the indirect distributional constraints, we have an uncertainty set using the newsvendor transaction data,

$$\mathcal{D}_T^o = \left\{\mathbb{P} \in \mathcal{P}(\Xi) : 1 - w_i^t/s_i - \Delta \le \mathbb{P}^i\left(\xi_i \le q_i^t\right) \le 1 - w_i^t/s_i + \Delta, \forall i \in [n], \forall t \in [T]\right\}. \tag{EC.14}$$

Table EC.3 The optimal and approximate value of the worst-case profit

	$\rho=0.3$			$\rho=0.6$			$\rho=0.9$		
(n,T)	Approx.	Optimal	Gap(%)	Approx.	Optimal	Gap(%)	Approx.	Optimal	Gap(%)
(5,10)	3780	3682	2.6	3711	3509	5.4	3675	3363	8.5
(5,20)	3916	3818	2.5	3865	3668	5.1	3804	3485	8.4
(5,30)	4008	3912	2.4	3971	3772	5.0	3922	3601	8.2
(10,10)	7067	6868	2.8	6992	6593	5.7	6921	6305	8.9
(10,20)	7596	7399	2.6	7484	7080	5.4	7386	6743	8.7
(10,30)	7990	7782	2.6	7893	7479	5.2	7828	7155	8.6
(15,10)	11466	11110	3.1	11343	10651	6.1	11176	10135	9.3
(15,20)	12398	12026	3.0	12236	11500	6.0	12124	10997	9.3
(15,30)	12865	12513	2.7	12690	11954	5.8	12494	11357	9.1

Note: ρ, correlation coefficient; Optimal: optimal value; Approx.: approximate value

Here, we integrate the quantile information with the empirical demand data to obtain an augmented center distribution of the Wasserstein ball of end-market demand. To solve the new center distribution with minimum Wasserstein distance, we can develop an equivalent linear programming reformulation via the LRDS method that is consistent with the approach in our original version.

$$\inf_{\boldsymbol{\beta}} \sum_{t=1}^{T}\sum_{m=1}^{M} \beta_{mt} \parallel \boldsymbol{\xi}_{mt}^{*} - \boldsymbol{\xi}^{t} \parallel_2^2 \tag{EC.15a}$$

$$\text{s.t.} \sum_{m=1}^{M} \beta_{mt} = 1/T, \forall t \in [T], \tag{EC.15b}$$

$$\sum_{t=1}^{T}\sum_{m=1}^{M} \beta_{mt} z_{im}\left(\boldsymbol{q}^{t'}\right) \le \frac{s_i - w_i^{t'}}{s_i} + \Delta, \forall i \in [n], t' \in [T], \tag{EC.15c}$$

$$\sum_{t=1}^{T}\sum_{m=1}^{M} \beta_{mt} z_{im}\left(\boldsymbol{q}^{t'}\right) \ge \frac{s_i - w_i^{t'}}{s_i} - \Delta, \forall i \in [n], t' \in [T], \tag{EC.15d}$$

$$\beta_{mt} \ge 0, \forall t \in [T], m \in [M], \tag{EC.15e}$$

Furthermore, we can develop an RWDRO pricing model with dual-source data and use the PCP method to solve the seller's worst-case profit. With (EC.14), we can obtain similar (relatively relaxed) upper and lower bounds. Specifically, for $i \in [n]$, we rearrange the past wholesale prices in descending order $s_i \ge w_i^1 > ... > w_i^t > ... > w_i^T > 0$, and rearrange the buyer's corresponding past order quantities in ascending order $0 \le q_i^1 < ... < q_i^t < ... < q_i^T < \overline{\xi}_i$. Given $\boldsymbol{w} = (w_i)_{i\in[n]}$, for any $i \in [n]$, let $\overline{w_i} = \min\{w_i^t : 1 - w_i/s_i \ge 1 - w_i^t/s_i + \Delta, t \in [T]\}$, and the corresponding buyer's past order under $\overline{w_i}$ is $\underline{q_{i,w}}$. Let $\underline{w_i} = \max\{w_i^t : 1 - w_i/s_i < 1 - w_i^t/s_i - \Delta, t \in [T]\}$, and the corresponding buyer's past order under $\underline{w_i}$ is $\overline{q_{i,w}}$. The past T order data partition region Ξ into $(T+1)^n$ subregions. And we also have that the buyer's worst-case order quantity exists only in a specific subregion $\times_{i=1}^{n}\left[\underline{q_{i,w}}, \overline{q_{i,w}}\right)$. Then, due to the same structure as the model in our original model, we can use the cutting-plane policy to search for the optimal solution within this subregion. When the products are independent, an equivalent SOCP reformulation can also be obtained.

With the above ingredients, we conducted a numerical analysis to analyze the impact of introducing additional tolerance on demand forecasting and out-of-sample performance through numerical

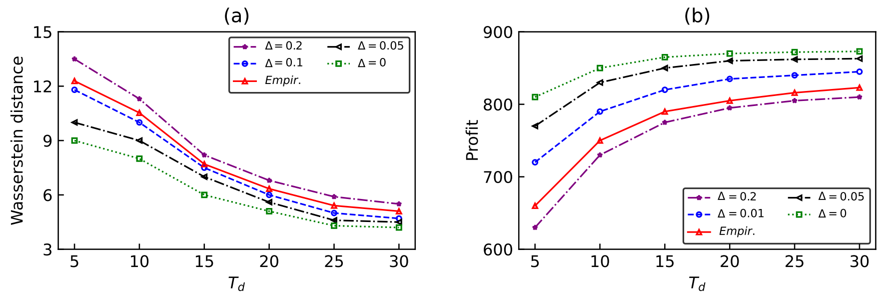

Figure EC.2 Wasserstein distance (a) and seller's profit (b) with different τ ($T_o = 5$)

experiments. As illustrated in Figure EC.2, we calculated the Wasserstein distance from the center distribution to the true distribution, as well as the seller's true profit with different values of $\varDelta$. As expected, when $\varDelta = 0$, the results align with the existing conclusions in our paper. As $\varDelta$ increases, the incremental value of incorporating historical order data for improving the accuracy of end-market demand forecasting gradually diminishes. When $\varDelta = 0.2$, the prediction accuracy and the seller's worst-case profit are even lower than using demand data alone. This is, the buyer's information advantage diminishes progressively as $\varDelta$ increases. Since the buyer's historical order records directly reflect his perception of demand, the above results suggest that suppliers should be cautious about incorporating historical purchase or transaction information into contract decisions when buyers also have a limited understanding of their end-market data.

EC.6. Data Sampling Rules in Numerical Experiments

For the case with a single product, the true distribution $\overline{\mathbb{P}}$ follows truncated normal distribution within $\Xi = \left[0, \overline{\xi}\right]$, with mean μ and standard deviation σ, denoted by $\overline{\mathbb{P}} \sim \mathcal{TN}(\mu, \sigma, \Xi)$. All historical demand points $\{\xi^t\}_{t\in[T_d]}$ are independently sampled from the distribution $\overline{\mathbb{P}}$. All historical wholesale prices $\{w^t\}_{t\in[T_o]}$ are uniformly sampled from the interval (c, s). The corresponding historical buyer orders $\{q^t\}_{t\in[T_o]}$ are optimal under the distribution $\overline{\mathbb{P}}$ $\left(\overline{\mathbb{P}}(\xi \le q^t) = 1 - w^t/s, \forall t \in [T_o]\right)$.

For the case with independent multi-products, the true distribution is $\overline{\mathbb{P}} = \times \overline{\mathbb{P}}^i$, where $\overline{\mathbb{P}}^i \sim \mathcal{TN}(\mu_i, \sigma_i, \Xi_i)$. For any $i \in [n]$, the historical demand points $\{\xi_i^t\}_{t\in[T_d]}$ are independently sampled from the distribution $\overline{\mathbb{P}}^i$. The historical wholesale prices $\{w_i^t\}_{t\in[T_o]}$ are uniformly sampled from the interval (c_i, s_i), and the historical orders $\{q_i^t\}_{t\in[T_o]}$ are optimal under $\overline{\mathbb{P}}^i$.

For the case with dependent multi-products, the true distribution $\overline{\mathbb{P}}$ follows multivariate truncated normal distribution within $\Xi = \left[0, \overline{\xi}\right]^n$, with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, denoted by $\overline{\mathbb{P}} \sim \mathcal{MTN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \Xi)$. All historical demand $\left\{\boldsymbol{\xi}^t\right\}_{t\in[T_d]}$ are independently sampled from the distribution $\overline{\mathbb{P}}$. For any $i \in [n]$, the historical prices $\{w_i^t\}_{t\in[T_o]}$ are uniformly sampled from interval (c_i, s_i), and corresponding historical orders $\{q_i^t\}_{t\in[T_o]}$ are optimal under the marginal distribution $\overline{\mathbb{P}}^i$.