

---

**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.**

## EC.1. Derivation of the Lagrangian Relaxation

LaBS heuristic in Miao et al. (2022) uses a hypothetical sell-back option at the end of the horizon which sells the remaining inventory in stores to the warehouse. The purpose is to have the optimal order-up-to level constant over time for each store after Lagrangian relaxation. While the reasoning is well documented in Miao et al. (2022), we add it here for the purpose of being self-contained. The sell-back adjusted dynamic program is written as

$$\begin{aligned}
\tilde{V}^* := \min_{\pi} & \sum_{t=1}^T \sum_{i=1}^N (\mathbb{E}[b_i(D_{i,t} - y_{i,t}^{\pi})^+ + h_i(y_{i,t}^{\pi} - D_{i,t})^+] + c'_i \mathbb{E}[y_{i,t}^{\pi} - I_{i,t}^{\pi}]) + wW & \text{(PI)} \\
& - \sum_{i=1}^N c'_i \mathbb{E}[(y_{i,T}^{\pi} - D_{i,T})^+] \\
\text{st} & \sum_{t=1}^T \sum_{i=1}^N (y_{i,t}^{\pi} - I_{i,t}^{\pi}) \leq W + \sum_{i=1}^N (y_{i,T}^{\pi} - D_{i,T})^+ \\
& I_{i,t+1}^{\pi} = (y_{i,t}^{\pi} - D_{i,t})^+ \quad t \in [T], i \in [N] \\
& y_{i,t}^{\pi} \geq I_{i,t}^{\pi} \quad t \in [T], i \in [N].
\end{aligned}$$

Since the constraint  $\sum_{t=1}^T \sum_{i=1}^N (y_{i,t}^{\pi} - I_{i,t}^{\pi}) \leq W + \sum_{i=1}^N (y_{i,T}^{\pi} - D_{i,T})^+$ , which has to be satisfied almost surely, links all the stores together, it is very difficult to solve (PI) directly. As a result, we first relax it to be satisfied in expectation, i.e.,

$$\sum_{t=1}^T \sum_{i=1}^N \mathbb{E}[(y_{i,t}^{\pi} - I_{i,t}^{\pi})] \leq W + \sum_{i=1}^N \mathbb{E}[(y_{i,T}^{\pi} - D_{i,T})^+].$$

Then we apply a Lagrangian relaxation on this constraint so that after relaxation, the problem becomes

$$\begin{aligned}
\tilde{V}^{\lambda} = \min_{\pi} & \sum_{t=1}^T \sum_{i=1}^N (\mathbb{E}[b_i(D_{i,t} - y_{i,t}^{\pi})^+ + h_i(y_{i,t}^{\pi} - D_{i,t})^+] + c'_i \mathbb{E}[y_{i,t}^{\pi} - I_{i,t}^{\pi}]) + wW & \text{(PS)} \\
& - \sum_{i=1}^N c'_i \mathbb{E}[(y_{i,T}^{\pi} - D_{i,T})^+] + \lambda \mathbb{E} \left[ \sum_{t=1}^T \sum_{i=1}^N (y_{i,t}^{\pi} - I_{i,t}^{\pi}) - W - \sum_{i=1}^N (y_{i,T}^{\pi} - D_{i,T})^+ \right] \\
\text{st} & I_{i,t+1}^{\pi} = (y_{i,t}^{\pi} - D_{i,t})^+ \quad t \in [T], i \in [N] \\
& y_{i,t}^{\pi} \geq I_{i,t}^{\pi} \quad t \in [T], i \in [N].
\end{aligned}$$

The purpose of this hypothetical sell-back option is because the relaxed problem (PS) has constant order-up-to level across the horizon. To see this, we note that after relaxation, we can separate the optimization for each store as

$$\begin{aligned}
\tilde{V}_i^{\lambda} := \min_{\pi} & \sum_{t=1}^T (\mathbb{E}[b_i(D_{i,t} - y_{i,t}^{\pi})^+ + (h_i - c_i(\lambda))(y_{i,t}^{\pi} - D_{i,t})^+] + c_i(\lambda) \mathbb{E}[y_{i,t}^{\pi}]) \\
\text{st} & I_{i,t+1}^{\pi} = (y_{i,t}^{\pi} - D_{i,t})^+ \quad t \in [T] & \text{(NV)} \\
& y_{i,t}^{\pi} \geq I_{i,t}^{\pi} \quad t \in [T],
\end{aligned}$$

which is a standard inventory control problem with i.i.d. demand over time. Therefore, a stationary order-up-to level  $y_i$  minimizing

$$\mathbb{E}[b_i(D_{i,1} - y)^+ + (h_i - c_i(\lambda))(y - D_{i,1})^+] + c_i(\lambda)\mathbb{E}[y]$$

suffices to solve for  $\tilde{V}_i^\lambda$ .

## EC.2. Technical Proofs in Section 4

LEMMA EC.1. *Given a fixed dual variable  $\lambda$ , and the constant  $\mathcal{C}_1 \geq \sqrt{3/2}(h_i + b_i - c_i)\sqrt{\log(T)}$ , with probability  $1 - O(T^{-2})$ , if Algorithm 1 runs during  $\nu$  periods, for any store  $i \in [N]$  the following hold.*

- i.  $\sum_{t \in \mathcal{T}_\nu} (C_i(y_{i,t}; \lambda) - C_i(y_i^*(\lambda); \lambda)) \leq \log_2(\nu \bar{D}) (\mathcal{C}_1^2 \underline{f}^{-1} h_i \log(\nu) + (h_i + b_i - c_i) + \bar{b} \bar{D})$  where  $\mathcal{T}_\nu$  is the set of time periods for Algorithm 1, and  $y_{i,t}$  is the actual base-stock level in time  $t$ .*
- ii.  $|y_{i,l} - y_i^*(\lambda)| \leq \frac{1}{\nu} + \frac{c_1 \sqrt{\lceil \log_2(\nu \bar{D}) \rceil}}{h_i \underline{f} \sqrt{\nu - \lceil \log_2(\nu \bar{D}) \rceil}}$  where  $y_{i,l}$  is the final output of the base-stock level.*
- iii.  $|\hat{d}_{i,l} - \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}]| \leq \left(\bar{D} + \frac{c_1}{h_i \underline{f}}\right) \sqrt{\frac{\log(NT) \lceil \log_2(\nu \bar{D}) \rceil}{\nu - \lceil \log_2(\nu \bar{D}) \rceil}} + \frac{1}{\nu}$  where  $\hat{d}_{i,l}$  is the final output of the estimated expected sales.*
- iv.  $|\sum_{t \in \mathcal{T}_\nu} Z_{i,t} - \nu \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}]| \leq K_1 \log(NT) \sqrt{\nu} + K_2$  where  $Z_{i,t}$  is the actual ordering in time  $t$  for store  $i$ ,  $K_1 = 1 + \sqrt{3}(h_i + b_i - c_i - \lambda)/(h_i \underline{f})$  and  $K_2 = \bar{D}(\bar{l} + 1)(\underline{D} + 1)/\underline{D}$ .*

Furthermore, the total number of extra periods to lower the inventory level to  $y_{i,l}$  does not exceed  $O(\log(\nu))$ .

*Proof of Lemma EC.1* Before we go into each part of this lemma, let us define

$$\mathcal{E}_1 := \{C'_i(y_{i,l}; \lambda) \in [g_l, \bar{g}_l] \text{ for all } l\},$$

which is the event that the gradient is included in the interval  $[g_l, \bar{g}_l]$  for all  $l$ . We shall show this event holds with probability at least  $1 - O(T^{-2})$ , and the rest of the proof shall be on this event. Note that the proof of parts *i.* and *ii.* mostly resembles Lemma 1 in Chen et al. (2020c), and we write the whole proof here for being self-contained. Note that in the rest of the proof, we shall assume  $\lambda < b_i - c_i$  because otherwise  $y_i^*(\lambda) = 0$  and all the results trivially hold. Also, for the last claim of this lemma, it holds because each loop has at most  $\bar{D}/\underline{D}$  periods of no ordering, which adds up to at most  $O(\log(\nu))$ .

$\hat{g}$  is an unbiased estimation of  $C'_i(y_{i,l}; \lambda)$ , and we have by Hoeffding's inequality (Hoeffding 1963)

$$P\left(C'_i(y_{i,l}; \lambda) \notin [\hat{g} - \zeta, \hat{g} + \zeta]\right) \leq 2 \exp\left(\frac{-2n_l \zeta^2}{(h_i + b_i - c_i)^2}\right),$$

for any  $\zeta > 0$ . Setting  $\zeta = \frac{\sqrt{3/2}(h_i + b_i - c_i)\sqrt{\log(T)}}{\sqrt{n_i}}$ , and having the constant  $\mathcal{C}_1 \geq \sqrt{3/2}(h_i + b_i - c_i)\sqrt{\log(T)}$  we have  $1 - O(T^{-3})$  probability for the gradient included in the confidence interval. Applying union bound for  $\nu \leq T$ , we have

$$P(\mathcal{E}_1) = P\left(C'_i(y_{i,l}; \lambda) \in [\underline{g}_l, \bar{g}_l] \forall l\right) \geq 1 - O(T^{-2}).$$

Therefore, when  $\underline{g}_l > 0$ , then  $C'_i(y_{i,l}; \lambda) > 0$  and when  $\bar{g}_l < 0$ ,  $C'_i(y_{i,l}; \lambda) < 0$ . This implies that  $y_i^*(\lambda)$  is always between  $L_l$  and  $R_l$ .

*Proof of Part i.* Let  $\mathcal{T}_l$  be the set of time periods during  $l$ . Suppose  $l < \lceil \log_2(\nu \bar{D}) \rceil$ , and consider  $l_0$ -th iteration in loop  $l$ . Given  $\mathcal{E}_1$ , we have that

$$|C'_i(y_{i,l}; \lambda)| \leq \mathcal{C}_1 \frac{1}{\sqrt{l_0 - 1}}. \quad (\text{EC.1})$$

Recall that according to Assumption 1 the probability density has  $0 < \underline{f} \leq f^i(d) \leq \bar{f} < \infty$  for all  $i \in [N]$  and  $\underline{f}, \bar{f} \in \mathbb{R}$ . Note that  $C''_i(y; \lambda) \geq (h_i + b_i - c_i - \lambda)\underline{f} \geq h_i \underline{f}$  for any order-up-to level  $y \in [\underline{D}, \bar{D}]$  and  $0 \leq \lambda < b_i - c_i$ . Thus, Taylor's theorem gives that

$$C_i(y_i^*(\lambda); \lambda) - C_i(y_{i,l}; \lambda) = C'_i(y_{i,l}; \lambda)(y_i^*(\lambda) - y_{i,l}) + C''_i(r; \lambda)(y_i^*(\lambda) - y_{i,l})^2 \leq 0,$$

where  $r \in \mathbb{R}$  is between  $y_i^*(\lambda)$  and  $y_{i,l}$ . Consequently,

$$C'_i(y_{i,l}; \lambda)(y_i^*(\lambda) - y_{i,l}) + h_i \underline{f} (y_i^*(\lambda) - y_{i,l})^2 \leq 0.$$

Rearranging the sides

$$h_i \underline{f} |y_i^*(\lambda) - y_{i,l}| \leq -C'_i(y_{i,l}; \lambda) \leq |C'_i(y_{i,l}; \lambda)| \leq \mathcal{C}_1 \frac{1}{\sqrt{l_0 - 1}}.$$

Therefore we have

$$|y_i^*(\lambda) - y_{i,l}| \leq \frac{\mathcal{C}_1}{h_i \underline{f} \sqrt{l_0 - 1}}. \quad (\text{EC.2})$$

For each loop  $l$ , let  $R_{\mathcal{T}_l} = \sum_{t \in \mathcal{T}_l} (C_i(y_{i,t}; \lambda) - C_i(y_i^*(\lambda); \lambda))$ .

**Case 1.**  $l < \lceil \log_2(\nu \bar{D}) \rceil$ : For any  $l_0$ -th time period in loop  $l$ , we have

$$\begin{aligned} C_i(y_{i,l}; \lambda) - C_i(y_i^*(\lambda); \lambda) &\leq C'_i(y_{i,l}; \lambda) |y_i^*(\lambda) - y_{i,l}| \leq \mathcal{C}_1 \frac{1}{\sqrt{l_0 - 1}} \cdot \frac{\mathcal{C}_1}{h_i \underline{f} \sqrt{l_0 - 1}} \\ &= \frac{\mathcal{C}_1^2}{h_i \underline{f} (l_0 - 1)}, \end{aligned}$$

where the first inequality follows from convexity of  $C_i(\cdot; \lambda)$  and the second one follows from (EC.20) and (EC.21).

Summing the cycle regret gives

$$R_{\tau_l} \leq \bar{b}\bar{D} + \sum_{l_0=2}^{|\tau_l|} \frac{C_1^2}{h_i \underline{f}(l_0 - 1)} \leq C_1^2 \underline{f}^{-1} h_i^{-1} \log(\nu) + \bar{b}\bar{D}.$$

**Case 2**  $l = \lceil \log_2(\nu \bar{D}) \rceil$ : Note that  $R_l - L_l \leq \bar{D}(\frac{1}{2})^{\lceil \log_2(\nu \bar{D}) \rceil} = \frac{1}{\nu}$ . Hence,  $|y_i^*(\lambda) - y_{i,l}| \leq 1/\nu$ . Then, by convexity,

$$C_i(y_{i,l}; \lambda) - C_i(y_i^*(\lambda); \lambda) \leq C'_i(y_{i,l}; \lambda) |y_i^*(\lambda) - y_{i,l}| \leq \frac{h_i + b_i - c_i}{\nu};$$

thus the summation over the cycle gives that

$$R_{\tau_l} \leq \frac{h_i + b_i - c_i}{\nu} n_l \leq h_i + b_i - c_i.$$

Combining the two cases, we have

$$\begin{aligned} \sum_{t=1}^{\nu} (C_i(y_{i,t}; \lambda) - C_i(y_i^*(\lambda); \lambda)) &\leq \sum_l R_{\tau_l} \\ &\leq \log_2(\nu \bar{D}) (C_1^2 \underline{f}^{-1} h_i \log(\nu) + (h_i + b_i - c_i) + \bar{b}\bar{D}), \end{aligned}$$

which gives the result.

*Proof of Part ii.* Let  $l^*$  be the cycle with largest  $n_l$ .

**Case 1.**  $l^* < \lceil \log_2(\nu \bar{D}) \rceil$ : Let  $\bar{l}$  be the latest cycle. We have that

$$n_{l^*} \geq \frac{\nu}{\bar{l}} \geq \frac{\nu}{\lceil \log_2(\nu \bar{D}) \rceil},$$

where the first inequality follows from the pigeonhole principle. Then, by inequality (EC.21), we have

$$|y_i^*(\lambda) - y_{i,l^*}| \leq \frac{C_1}{h_i \underline{f} \sqrt{n_{l^*} - 1}} \leq \frac{C_1 \sqrt{\lceil \log_2(\nu \bar{D}) \rceil}}{h_i \underline{f} \sqrt{\nu - \lceil \log_2(\nu \bar{D}) \rceil}}.$$

**Case 2.**  $l^* = \lceil \log_2(\nu \bar{D}) \rceil$ : Then we have by binary search,  $|y_i^*(\lambda) - y_{i,l^*}| \leq 1/\nu$ .

Combining the two cases, we have

$$|y_{i,l^*}(\lambda) - y_i^*(\lambda)| \leq \frac{1}{\nu} + \frac{C_1 \sqrt{\lceil \log_2(\nu \bar{D}) \rceil}}{h_i \underline{f} \sqrt{\nu - \lceil \log_2(\nu \bar{D}) \rceil}}.$$

*Proof of Part iii.* Let loop  $l^*$  be the one with the largest  $n_l$ .  $\hat{d}_{i,l^*}$  is defined as

$$\hat{d}_{i,l^*} = \frac{\sum_{t \in \tau_{l^*}} \min\{y_{i,l^*}, D_{i,t}\}}{n_{l^*}}.$$

According to Hoeffding's inequality, we have that with probability at least  $1 - O(N^{-1}T^{-2})$ ,

$$\left| \hat{d}_{i,l^*} - \mathbb{E}[\min\{y_{i,l^*}, D_{i,1}\}] \right| \leq \bar{D} \sqrt{\frac{\log(NT)}{n_{l^*}}}.$$

Then by part *ii.*, we have

$$\begin{aligned} & \left| \hat{d}_{i,l^*} - \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}] \right| \\ & \leq \left| \hat{d}_{i,l^*} - \mathbb{E}[\min\{y_{i,l^*}, D_{i,1}\}] \right| + |\mathbb{E}[\min\{y_{i,l^*}, D_{i,1}\}] - \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}]| \\ & \leq \left| \hat{d}_{i,l^*} - \mathbb{E}[\min\{y_{i,l^*}, D_{i,1}\}] \right| + |y_{i,l^*} - y_i^*(\lambda)| \\ & \leq \bar{D} \sqrt{\frac{\log(NT)}{n_{l^*}}} + \frac{1}{\nu} + \frac{\mathcal{C}_1 \sqrt{\lceil \log_2(\nu \bar{D}) \rceil}}{h_i \underline{f} \sqrt{\nu - \lceil \log_2(\nu \bar{D}) \rceil}}. \end{aligned}$$

Since  $n_{l^*}$  is the largest among all  $n_l$ , and  $l \leq \lceil \log_2(\nu \bar{D}) \rceil$ , we must have  $n_{l^*} \geq \nu / \lceil \log_2(\nu \bar{D}) \rceil$ , implying that

$$\left| \hat{d}_{i,l^*} - \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}] \right| \leq \left( \bar{D} + \frac{\mathcal{C}_1}{h_i \underline{f}} \right) \sqrt{\frac{\log(NT) \lceil \log_2(\nu \bar{D}) \rceil}{\nu - \lceil \log_2(\nu \bar{D}) \rceil}} + \frac{1}{\nu}.$$

*Proof for Part iv.* We have

$$\begin{aligned} \sum_{t \in \mathcal{T}_\nu} Z_{i,t} &= \sum_l \sum_{t \in \mathcal{T}_l} Z_{i,t} \\ &\leq \sum_l \sum_{t \in \mathcal{T}_l} \min\{y_{i,t}, D_{i,t}\} + \bar{D}(\bar{l} + 1). \end{aligned}$$

where  $\bar{l}$  is the index of the last cycle, and the last inequality is due to possible extra ordering at the beginning of each cycle  $l$ . Similarly, we have

$$\sum_{t \in \mathcal{T}_\nu} Z_{i,t} \geq \sum_l \sum_{t \in \mathcal{T}_l} \min\{y_{i,t}, D_{i,t}\} - \bar{D}/\underline{D}(\bar{l} + 1),$$

where  $\bar{D}/\underline{D}(\bar{l} + 1)$  is from possible no-order periods in each loop.

Then

$$\begin{aligned}
& \left| \sum_{t \in \mathcal{T}_\nu} Z_{i,t} - \nu \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}] \right| \\
& \leq \left| \sum_l \sum_{t \in \mathcal{T}_l} \min\{y_{i,t}, D_{i,t}\} - \sum_{l=0}^{\bar{l}} n_l \mathbb{E}[\min\{y_{i,l}, D_{i,1}\}] \right| \\
& \quad + \sum_{l=0}^{\bar{l}} n_l |\mathbb{E}[\min\{y_{i,l}, D_{i,1}\}] - \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}]| + \bar{D}(\bar{l}+1) \frac{\underline{D}+1}{\underline{D}} \\
& \leq \sum_{l=0}^{\bar{l}} \left( \sqrt{\log(NT)n_l} \right) + \sum_{l=0}^{\bar{l}} n_l |\mathbb{E}[\min\{y_{i,l}, D_{i,1}\}] - \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}]| + \bar{D}(\bar{l}+1) \frac{\underline{D}+1}{\underline{D}} \\
& \leq \sqrt{\log(NT)\nu} + \frac{\nu \mathcal{C}_1 \sqrt{\lceil \log_2(\nu \bar{D}) \rceil}}{h_i \underline{f} \sqrt{\nu - \lceil \log_2(\nu \bar{D}) \rceil}} + \bar{D}(\bar{l}+1) \frac{\underline{D}+1}{\underline{D}} \\
& \leq \sqrt{\log(NT)\nu} + \frac{\mathcal{C}_1 \sqrt{2 \lceil \log_2(\nu \bar{D}) \rceil}}{h_i \underline{f}} \sqrt{\nu} + \bar{D}(\bar{l}+1) \frac{\underline{D}+1}{\underline{D}} \\
& \leq K_1 \log(NT) \sqrt{\nu} + K_2,
\end{aligned}$$

where the second inequality is satisfied with probability at least  $1 - O(N^{-1}T^{-2})$  using Hoeffding's inequality, the third inequality holds according to a similar argument as in Part *ii* and the constants  $K_1 = 1 + \frac{\sqrt{3}(h_i + b_i - c_i - \lambda)}{h_i \underline{f}}$  and  $K_2 = \bar{D}(\bar{l}+1) \frac{\underline{D}+1}{\underline{D}}$ .

□

In the next lemma, we prove an important convergence result with respect to  $\hat{\lambda}_\tau$  in our algorithm DBS.

LEMMA EC.2. *Let*

$$\begin{aligned}
\mathcal{C}_0 & \geq \max(4/\beta^2, 2 \lceil \log_2(T\bar{D}) \rceil) \\
\mathcal{C}_2 & \geq \left( \bar{D} + \frac{\mathcal{C}_1}{h_i \underline{f}} \right) \sqrt{2 \log(NT) \lceil \log_2(T\bar{D}) \rceil} \\
\mathcal{C}_3 & \geq \max \left( \sqrt{\mathcal{C}_0} \min_i (b_i - c_i), 2\mathcal{C}_2 / \min_i \frac{h_i^2}{(h_i + b_i - c_i)^3 \underline{f}} \right) \\
& \beta \in (1, 4],
\end{aligned}$$

then with probability at least  $1 - O(T^{-1})$ , we have

$$|\hat{\lambda}_\tau - \lambda^*| \leq \mathcal{C}_3 \beta^{-\tau/2+3/2},$$

for all  $\tau$ .

*Proof of Lemma EC.2* Define the event

$$\mathcal{E}_2 := \left\{ \left| \sum_{i \in [N]} \hat{d}_{i,\tau} - \sum_{i \in [N]} \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] \right| \leq \mathcal{C}_2 N / \sqrt{\nu_\tau}, \forall \tau \right\}.$$

Note that

$$\left| \sum_{i \in [N]} \hat{d}_{i;\tau} - \sum_{i \in [N]} \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] \right| \leq \sum_{i \in [N]} \left| \hat{d}_{i;\tau} - \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] \right| \leq \mathcal{C}_2 N / \sqrt{\nu_\tau},$$

where the last inequality holds with probability at least  $1 - O(T^{-1})$  according to Lemma EC.1.(iii.) by union bound over  $\tau$  and the fact that  $\nu_0 = \mathcal{C}_0 \geq 2\lceil \log_2(T\bar{D}) \rceil$  and  $\mathcal{C}_2 \geq \left( \bar{D} + \frac{\mathcal{C}_1}{h_{i,f}} \right) \sqrt{2\log(NT)\lceil \log_2(T\bar{D}) \rceil}$ . Thus event  $\mathcal{E}_2$  holds with probability at least  $1 - O(T^{-1})$ . Our argument will be conditioned on event  $\mathcal{E}_2$  holds.

By a bit abuse of notation, let us write  $\tilde{V}^\lambda$  as its normalized version (by  $T$ , i.e.,  $\tilde{V}^\lambda/T$ ). For notation convenience, let  $\tilde{V}'(\lambda) := \sum_i \mathbb{E}[\min(y_i^*(\lambda), D_{i,1})] - W/T = \frac{d\tilde{V}^\lambda}{d\lambda}$ . With  $\mathcal{E}_2$  holds, this shows that  $\tilde{V}'(\hat{\lambda}_\tau) = \sum_{i \in [N]} \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] - W/T \in [\underline{v}_\tau, \bar{v}_\tau]$  for all  $\tau$ . Moreover, we have, by inverse function theorem and Assumption 1 and Assumption 2,

$$\begin{aligned} \frac{d^2 \tilde{V}^\lambda}{d\lambda^2} &= - \sum_{i \in [N]} \frac{h_i}{(h_i + b_i - c_i - \lambda)^2 f^i(y_i^*(\lambda))} (1 - F^i(y_i^*(\lambda))) \\ &= - \sum_{i \in [N]} \frac{h_i^2}{(h_i + b_i - c_i - \lambda)^3 f^i(y_i^*(\lambda))} \\ &\leq - \sum_{i \in [N]} \frac{h_i^2}{(h_i + b_i - c_i)^3 \bar{f}} \leq -N \min_{i \in [N]} \frac{h_i^2}{(h_i + b_i - c_i)^3 \bar{f}}. \end{aligned}$$

Let us extend  $\tilde{V}'(\lambda)$  to the domain of negative  $\lambda$ . The way we do it is by extending  $\tilde{V}'(\lambda)$  linearly with the rate of  $\frac{d^2 \tilde{V}^0}{d\lambda^2}$  so that  $\tilde{V}'(\lambda)$  is differentiable after extension and we still have the same upper bound of  $\frac{d^2 \tilde{V}^\lambda}{d\lambda^2}$  for all  $\lambda \leq \min_i(b_i - c_i)$ . Note that since  $\frac{d^2 \tilde{V}^0}{d\lambda^2} < 0$ , we have  $\tilde{V}'(\lambda) \rightarrow \infty$  if  $\lambda \rightarrow -\infty$ , which implies, together with Assumption 2, that there always exists  $\tilde{\lambda} \in \mathbb{R}$  such that  $\tilde{V}'(\tilde{\lambda}) = 0$ . Obviously, if  $\tilde{\lambda} < 0$ , we have  $\lambda^* = 0$  and otherwise  $\lambda^* = \tilde{\lambda}$ .

For the rest of the proof, we **claim** that  $U_\tau - L_\tau \leq \beta^{3/2} \mathcal{C}_3 \beta^{-\tau/2}$  and  $\lambda^* \in [L_\tau, U_\tau]$  for all  $\tau$ . Obviously, if this claim is true, we immediately have our result. We prove this by induction. This statement is obviously true for  $\tau = 0$  by assuming  $\beta^{3/2} \mathcal{C}_3 \geq \min_i(b_i - c_i)$ . For  $\tau = 1$ , we have one of the following conditions.

If line 7 happens, we have  $U_\tau, L_\tau$  unchanged. Then just take  $\beta^{3/2} \mathcal{C}_3 \beta^{-1/2} \geq \min_i(b_i - c_i)$  and we are good. If line 10 happens, we immediately have  $\lambda^* = 0$  and  $U_\tau = L_\tau = 0$ , so we are good. If line 13 happens, since  $|\hat{\lambda}_0 - \lambda^*| = |\lambda^*| \leq \min_i(b_i - c_i) \leq \mathcal{C}_3 / \sqrt{\mathcal{C}_0}$  where the last inequality is because  $\mathcal{C}_3 \geq \sqrt{\mathcal{C}_0} \min_i(b_i - c_i)$ . Thus taking  $U_1 = \text{Proj}_{[0, \min_i(b_i - c_i)]}(\hat{\lambda}_0 + \mathcal{C}_3 / \sqrt{\mathcal{C}_0})$  and  $L_1 = \text{Proj}_{[0, \min_i(b_i - c_i)]}(\hat{\lambda}_0 - \mathcal{C}_3 / \sqrt{\mathcal{C}_0})$  has  $\lambda^* \in [L_1, U_1]$ , and

$$U_1 - L_1 \leq 2\mathcal{C}_3 / \sqrt{\mathcal{C}_0} \leq \beta^{3/2} \mathcal{C}_3 \beta^{-1/2},$$

because  $\mathcal{C}_0 \geq 4/\beta^2$ , which again proves the claim.

For the inductive step, suppose it holds for  $\tau$ ; let us consider  $\tau + 1$ . The procedure is almost the same as for  $\tau = 1$ .

If condition line 7 or 10 is satisfied for  $\tau$ , then by algorithm design, we have

$$U_{\tau+1} - L_{\tau+1} = (U_\tau - L_\tau)/2 \leq \beta^{3/2} \mathcal{C}_3 \beta^{-\tau/2} / 2 \leq \beta^{3/2} \mathcal{C}_3 \beta^{-(\tau+1)/2},$$

where the first inequality is from the inductive hypothesis, and the last inequality is because  $\beta \leq 4$ .

Moreover, since  $\tilde{V}'(\hat{\lambda}_\tau) = \sum_{i \in [N]} \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] - W/T \in [\underline{v}_\tau, \bar{v}_\tau]$ , line 7 (10) implies that  $\hat{\lambda}_\tau \leq \lambda^*$  ( $\hat{\lambda}_\tau \geq \lambda^*$ ), which shows that  $\lambda^* \in [L_{\tau+1}, U_{\tau+1}]$ . Thus our claim is proved.

If condition line 13 is satisfied for  $\tau$ , this implies that  $0, \tilde{V}'(\hat{\lambda}_\tau) \in [\underline{v}_\tau, \bar{v}_\tau]$  and thus  $|\tilde{V}'(\hat{\lambda}_\tau)| \leq 2\mathcal{C}_2 N / \sqrt{\nu_\tau}$ . Note that by Taylor's theorem on  $\tilde{V}'(\lambda)$ , we have

$$\tilde{V}'(\hat{\lambda}_\tau) = \tilde{V}'(\hat{\lambda}_\tau) - \tilde{V}'(\tilde{\lambda}) = \frac{d^2 \tilde{V}^{\tilde{\lambda}}}{d\lambda^2} (\hat{\lambda}_\tau - \tilde{\lambda}),$$

where  $\tilde{\lambda}$  is between  $\hat{\lambda}_\tau$  and  $\tilde{\lambda}$ . This implies that

$$\begin{aligned} |\hat{\lambda}_\tau - \lambda^*| &\leq |\hat{\lambda}_\tau - \tilde{\lambda}| \\ &\leq \frac{1}{N \min_i \frac{h_i^2}{(h_i + b_i - c_i)^{3\bar{f}}}} \left| \tilde{V}'(\hat{\lambda}_\tau) \right| \\ &\leq \frac{2\mathcal{C}_2 / \sqrt{\nu_\tau}}{\min_i \frac{h_i^2}{(h_i + b_i - c_i)^{3\bar{f}}}} \\ &\leq \mathcal{C}_3 / \sqrt{\nu_\tau}, \end{aligned}$$

where the last inequality is because  $\mathcal{C}_3 \geq 2\mathcal{C}_2 / \min_i \frac{h_i^2}{(h_i + b_i - c_i)^{3\bar{f}}}$ . Thus taking  $U_{\tau+1} = \text{Proj}_{[0, \min_i (b_i - c_i)]}(\hat{\lambda}_\tau + \mathcal{C}_3 \beta^{-\tau/2} / \sqrt{\mathcal{C}_0})$  and  $L_{\tau+1} = \text{Proj}_{[0, \min_i (b_i - c_i)]}(\hat{\lambda}_\tau - \mathcal{C}_3 \beta^{-\tau/2} / \sqrt{\mathcal{C}_0})$  has  $\lambda^* \in [L_{\tau+1}, U_{\tau+1}]$ , and

$$U_{\tau+1} - L_{\tau+1} \leq 2\mathcal{C}_3 \beta^{-\tau/2} / \sqrt{\mathcal{C}_0} \leq \beta^{3/2} \mathcal{C}_3 \beta^{-(\tau+1)/2},$$

because  $\mathcal{C}_0 \geq 4/\beta^2$ , which again proves the claim.

□

**A note on comparison between zeroth- and first-order methods:** Note that our DBS algorithm can be considered as a first-order method as we estimate the first-order derivative of value function at different  $\lambda$  values. Although one can devise an equivalent zeroth-order method that does not require derivative information, zeroth-order methods can lead to performance degradation compared to first-order methods. More specifically, even though in a zeroth-order method we can evaluate the value of  $\tilde{V}^\lambda$  at different  $\lambda$  without knowing its derivative, we need to utilize an algorithm to shrink the search range for  $\lambda^*$ . For example, as commonly used in the literature, a trisection search (see, e.g., Lei et al. 2014 in which it is named as double-bisection search) would lead to similar theoretical results as in our approach. However, when we have access to the (noisy)

first-order information, we can apply the bisection search which shrinks the search range of  $\lambda^*$  in a rate of  $O(2^{-\tau})$  for each iteration  $\tau$ , as opposed to the shrinking rate of  $O((2/3)^\tau)$  for trisection search. As a result, not only the length of each iteration  $\tau$  is shorter in our algorithm (as we only need to evaluate one  $\lambda_\tau$  instead of two for trisection search), but also the convergence to  $\lambda^*$  is faster. This could make a significant difference in empirical performance.

### EC.3. Sufficient Conditions and Relaxing Assumption 2

LEMMA EC.3. *Suppose one of the following conditions is satisfied.*

1. *All stores have homogeneous cost structure, i.e.,  $b_i = b$  and  $c_i = c$  for all  $i \in [N]$ .*
2. *The warehouse inventory is more than the expected demand, i.e.,  $W > \sum_{i=1}^N \sum_{t=1}^T E[D_{i,t}]$ .*

*Then, Assumption 2 is guaranteed to hold true.*

*Proof of Lemma EC.3* We have the complementary slackness conditions proved by Lemma 5 in Miao et al. (2022) imply that  $\lambda^*$  must satisfy  $\lambda^* \cdot \left( \sum_{t=1}^T \sum_{i=1}^N \mathbb{E}[\min(y_i^*(\lambda^*), D_{i,t})] - W \right) = 0$ ,  $\lambda^* \geq 0$ , and  $\sum_{t=1}^T \sum_{i=1}^N \mathbb{E}[\min(y_i^*(\lambda^*), D_{i,t})] \leq W$ .

We first show that the optimal  $\lambda^*$  (i.e., the maximizer of  $V_\lambda$  in the set  $\lambda \geq 0$ ) must satisfy  $\lambda^* \leq \max_{i \in [N]}(b_i - c_i)$ , otherwise, the system would not be worth operating. Assume the opposite, that is  $\lambda^* > \max_{i \in [N]}(b_i - c_i)$ . Then, we have for any  $i$ ,  $y_i^*(\lambda^*)$ , which is a minimizer of  $C'_i(y_i; \lambda)$ , must be equal to 0 because

$$C'_i(0; \lambda^*) = -(b_i - c_i(\lambda^*)) = \lambda^* - (b_i - c_i) > 0.$$

Then, if we use the complementary slackness condition

$$\lambda^* \cdot \left( \sum_{t=1}^T \sum_{i=1}^N \mathbb{E}[\min(y_i^*(\lambda^*), D_{i,t})] - W \right) = \lambda^* \cdot (-W) = 0,$$

we get  $\lambda^* = 0$  which contradicts with  $\lambda^* > \max_{i \in [N]}(b_i - c_i)$ . Therefore,  $\lambda^* \leq \max_{i \in [N]}(b_i - c_i)$  holds.

Now, if condition 1 is satisfied, then, we have  $\max_{i \in [N]}(b_i - c_i) = \min_{i \in [N]}(b_i - c_i)$ . Therefore, this implies that  $\lambda^* \leq \max_{i \in [N]}(b_i - c_i) = \min_{i \in [N]}(b_i - c_i)$ , which in turn guarantees that Assumption 2 holds true.

Suppose that condition 2 is satisfied. Then, it implies that  $W > \sum_{t=1}^T \sum_{i=1}^N \mathbb{E}[\min(y_i, D_{i,t})]$  for any  $y_i \geq 0$ . Based on the complementary slackness condition  $\lambda^* \cdot \left( \sum_{t=1}^T \sum_{i=1}^N \mathbb{E}[\min(y_i^*(\lambda^*), D_{i,t})] - W \right) = 0$ , we get  $\lambda^* = 0$ , which in turn guarantees that the Assumption 2 holds true.

□

Now, we will analyze the scenario when the Assumption 2 is not met.

LEMMA EC.4. *Let*

$$\begin{aligned}\mathcal{C}_0 &\geq \max(4/\beta^2, 2\lceil \log_2(T\bar{D}) \rceil) \\ \mathcal{C}_2 &\geq \left(\bar{D} + \frac{\mathcal{C}_1}{h_i \underline{f}}\right) \sqrt{2\log(NT)\lceil \log_2(T\bar{D}) \rceil} \\ \mathcal{C}_3 &\geq \max\left(\sqrt{\mathcal{C}_0} \min_i(b_i - c_i), 2\mathcal{C}_2 / \min_i \frac{h_i^2}{(h_i + b_i - c_i)^3 \underline{f}}\right) \\ \beta &\in (1, 2] \\ \tau_0 &:= \left\lceil \frac{2\log(N\mathcal{C}_3\beta^{3/2} / \min\{\lambda^* - b'_k, b'_{k+1} - \lambda^*\})}{\log(\beta)} \right\rceil,\end{aligned}$$

then with probability at least  $1 - O(T^{-1})$ , we have

$$|\hat{\lambda}_\tau - \lambda^*| \leq \tilde{O}(N\beta^{-\tau/2}),$$

for  $\tau < \tau_0$  and

$$|\hat{\lambda}_\tau - \lambda^*| \leq \tilde{O}(\max\{\beta^{-(\tau-\tau_0)}, \beta^{-\tau/2}\}),$$

for  $\tau \geq \tau_0$ .

*Proof of Lemma EC.4* Define the event

$$\mathcal{E}_2 := \left\{ \left| \sum_{i \in [N]} \hat{d}_{i;\tau} - \sum_{i \in [N]} \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] \right| \leq \mathcal{C}_2 N_\tau / \sqrt{\nu_\tau}, \forall \tau \right\}.$$

Note that

$$\left| \sum_{i \in [N]} \hat{d}_{i;\tau} - \sum_{i \in [N]} \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] \right| \leq \sum_{i \in [N]} \left| \hat{d}_{i;\tau} - \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] \right| \leq \mathcal{C}_2 N_\tau / \sqrt{\nu_\tau},$$

where the last inequality holds with probability at least  $1 - O(T^{-1})$  according to Lemma EC.1.(iii.) by union bound over  $\tau$  and the fact that  $\nu_0 = \mathcal{C}_0 \geq 2\lceil \log_2(T\bar{D}) \rceil$  and  $\mathcal{C}_2 \geq \left(\bar{D} + \frac{\mathcal{C}_1}{h_i \underline{f}}\right) \sqrt{2\log(NT)\lceil \log_2(T\bar{D}) \rceil}$ . Also, note that if some  $i$  has  $c_i + \hat{\lambda}_\tau \geq b_i$ , we must have  $\hat{d}_{i;\tau} = y_i^*(\hat{\lambda}_\tau) = 0$ , thus  $|\hat{d}_{i;\tau} - \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}]| = 0$ . Then event  $\mathcal{E}_2$  holds with probability at least  $1 - O(T^{-1})$ . Our argument will be conditioned on event  $\mathcal{E}_2$  holds.

By a bit abuse of notation, let us write  $\tilde{V}^\lambda$  as its normalized version (by  $T$ , i.e.,  $\tilde{V}^\lambda/T$ ). For notation convenience, let  $\tilde{V}'(\lambda) := \sum_i \mathbb{E}[\min(y_i^*(\lambda), D_{i,1})] - W/T = \frac{d\tilde{V}^\lambda}{d\lambda}$ . With  $\mathcal{E}_2$  holds, this shows that  $\tilde{V}'(\hat{\lambda}_\tau) = \sum_{i \in [N]} \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] - W/T \in [\underline{v}_\tau, \bar{v}_\tau]$  for all  $\tau$ .

We then extend  $\tilde{V}'(\lambda)$  to the domain of negative  $\lambda$  similarly as in Lemma EC.2 and define  $\tilde{\lambda}$  as the zero point of  $\tilde{V}'(\lambda)$ . First, we have that for  $\tau < \tau_0$ ,  $U_\tau - L_\tau \leq N\beta^{3/2}\mathcal{C}_3\beta^{-\tau/2}$  and  $\lambda^* \in [L_\tau, U_\tau]$ . This can be proved by induction following similar proof as in Lemma EC.2. The only difference is

that in the inductive step, if condition line 13 is satisfied for  $\tau$ , this implies that  $0, \tilde{V}'(\hat{\lambda}_\tau) \in [\underline{v}_\tau, \bar{v}_\tau]$  and  $|\tilde{V}'(\hat{\lambda}_\tau)| \leq 2\mathcal{C}_2 N_\tau / \sqrt{\nu_\tau}$ . Note that by Taylor's theorem on  $\tilde{V}'(\lambda)$ , we have

$$\tilde{V}'(\hat{\lambda}_\tau) = \tilde{V}'(\hat{\lambda}_\tau) - \tilde{V}'(\tilde{\lambda}) = \frac{d^2 \tilde{V}^{\tilde{\lambda}}}{d\lambda^2} (\hat{\lambda}_\tau - \tilde{\lambda}),$$

where  $\tilde{\lambda}$  is between  $\hat{\lambda}_\tau$  and  $\tilde{\lambda}$ . However, when  $\tau < \tau_0$ , we cannot guarantee  $\hat{\lambda}_\tau$  is very close to  $\tilde{\lambda}$ . Therefore, we can only have

$$\left| \frac{d^2 \tilde{V}^{\tilde{\lambda}}}{d\lambda^2} \right| \geq \min_{i \in [N]} \frac{h_i^2}{(h_i + b_i - c_i)^3 \bar{f}},$$

instead of greater than  $N_\tau \min_{i \in [N]} \frac{h_i^2}{(h_i + b_i - c_i)^3 \bar{f}}$ . Therefore, this implies

$$\begin{aligned} |\hat{\lambda}_\tau - \lambda^*| &\leq |\hat{\lambda}_\tau - \tilde{\lambda}| \\ &\leq \frac{1}{\min_i \frac{h_i^2}{(h_i + b_i - c_i)^3 \bar{f}}} \left| \tilde{V}'(\hat{\lambda}_\tau) \right| \\ &\leq \frac{2\mathcal{C}_2 N_\tau / \sqrt{\nu_\tau}}{\min_i \frac{h_i^2}{(h_i + b_i - c_i)^3 \bar{f}}} \\ &\leq \mathcal{C}_3 N_\tau / \sqrt{\nu_\tau}, \end{aligned}$$

where the last inequality is because  $\mathcal{C}_3 \geq 2\mathcal{C}_2 / \min_i \frac{h_i^2}{(h_i + b_i - c_i)^3 \bar{f}}$ . Thus taking  $U_{\tau+1} = \text{Proj}_{[0, \min_i (b_i - c_i)]}(\hat{\lambda}_\tau + \mathcal{C}_3 N_\tau \beta^{-\tau/2} / \sqrt{\mathcal{C}_0})$  and  $L_{\tau+1} = \text{Proj}_{[0, \min_i (b_i - c_i)]}(\hat{\lambda}_\tau - \mathcal{C}_3 N_\tau \beta^{-\tau/2} / \sqrt{\mathcal{C}_0})$  has  $\lambda^* \in [L_{\tau+1}, U_{\tau+1}]$ , and

$$U_{\tau+1} - L_{\tau+1} \leq 2\mathcal{C}_3 N_\tau \beta^{-\tau/2} / \sqrt{\mathcal{C}_0} \leq N \beta^{3/2} \mathcal{C}_3 \beta^{-(\tau+1)/2},$$

because  $\mathcal{C}_0 \geq 4/\beta^2$ , which proves the claim.

For  $\tau \geq \tau_0$ , by definition we have  $N \beta^{3/2} \mathcal{C}_3 \beta^{-\tau_0/2} \leq \min\{\lambda^* - b'_k, b'_{k+1} - \lambda^*\}$ , implying  $|\hat{\lambda}_\tau - \lambda^*| \leq \min\{\lambda^* - b'_k, b'_{k+1} - \lambda^*\}$  for all  $\tau \geq \tau_0$ . Therefore, we have  $N_\tau = N^* := |\{i \in [N] : c_i + \lambda^* < b_i\}|$ , and

$$\left| \frac{d^2 \tilde{V}^{\tilde{\lambda}}}{d\lambda^2} \right| \geq N_\tau \min_{i \in [N]} \frac{h_i^2}{(h_i + b_i - c_i)^3 \bar{f}}. \quad (\text{EC.3})$$

We claim that we must have for all  $\tau \geq \tau_0$ ,  $U_\tau - L_\tau \leq \beta^{3/2} \mathcal{C}_3 \max\{\beta^{-(\tau-\tau_0)}, \beta^{-\tau/2}\}$  and  $\lambda^* \in [U_\tau - L_\tau]$  thus completing the proof.

Again we prove by induction, where the base case is by definition of  $\tau_0$ . For the inductive step, if condition line 7 or 10 is satisfied for  $\tau$ , then by algorithm design, we have

$$U_{\tau+1} - L_{\tau+1} = (U_\tau - L_\tau) / 2 \leq \beta^{3/2} \mathcal{C}_3 \max\{\beta^{-(\tau-\tau_0)}, \beta^{-\tau/2}\} / 2 \leq \beta^{3/2} \mathcal{C}_3 \max\{\beta^{-(\tau+1-\tau_0)}, \beta^{-(\tau+1)/2}\},$$

where the second inequality is because  $\beta \leq 2$ . If line 13 holds, because of (EC.3), following the same procedure as Lemma EC.2 we will have

$$U_{\tau+1} - L_{\tau+1} \leq 2\mathcal{C}_3 \beta^{-\tau/2} / \sqrt{\mathcal{C}_0} \leq \beta^{3/2} \mathcal{C}_3 \beta^{-(\tau+1)/2},$$

because  $\mathcal{C}_0 \geq 4/\beta^2$ , and we are done.  $\square$

With Lemma EC.4, we are ready to prove the regret upper bound  $\tilde{O}(N\sqrt{T} + N^3)$ . The proof mostly resembles that of Theorem 1, so we neglect it for brevity. The only difference is that we need to analyze each difference before and after loop  $\tau_0$ . In particular, the regret before  $\tau_0$  is at most  $\tilde{O}(N\beta^{\tau_0}) \leq \tilde{O}(N^3)$  by definition of  $\tau_0$ , and the regret after  $\tau_0$  is at most  $\tilde{O}(N \sum_{\tau} \max\{\beta^{-(\tau-\tau_0)}, \beta^{-\tau/2}\} \beta^{\tau}) \leq \tilde{O}(N\sqrt{T} + N^3)$ .

#### EC.4. Technical Details for Uncensored Demand

In this section, we present the algorithm and its performance when there is no demand censoring. We first present a Lagrangian relaxation based benchmark, which is used for bounding the regret as well as developing the algorithm. Then our algorithm based on sample average approximation (SAA) is presented together with its theoretical performance. It shall be noted that besides approximation results which may be of independent interest, this section also shows the major differences and challenges of online optimization in OWMS system compared with single-store inventory control literature (see, e.g., Levi et al. 2007, Huh and Rusmevichientong 2009).

**The algorithm.** Without demand censoring, our algorithm, named EDAF, is based on sample average approximation. Briefly, at the beginning of time period  $t \geq 2$ , we have collected all the demand realizations  $D_{i,s}$  for all  $i \in [N]$  and  $s \leq t-1$  with full observation. Using these data, we can construct an empirical estimation  $\hat{F}_t^i(\cdot)$  of the demand distribution  $F^i(\cdot)$ . More specifically, we have

$$\hat{F}_t^i(d) := \frac{\sum_{s=1}^{t-1} \mathbf{1}(D_{i,s} \leq d)}{t-1}.$$

According to Dvoretzky–Kiefer–Wolfowitz inequality (see Massart 1990), as  $t$  grows,  $\hat{F}_t^i(\cdot)$  becomes more accurate (see Lemma 1).

Similarly, we have the corresponding empirical probability density function as

$$\hat{f}^i(d) = \frac{\sum_{s=1}^{t-1} \mathbf{1}(D_{i,s} = d)}{t-1}.$$

Once we have this empirical demand distribution, we are able to estimate  $y_i^*(\lambda)$  for any  $\lambda$ , which is defined as

$$\hat{y}_{i,t}(\lambda) := \inf\{y : \hat{F}_t^i(y) \geq \kappa_i(\lambda)\}. \quad (\text{EC.4})$$

Then for  $\lambda^*$ , according to Lemma 10 in Miao et al. (2022),  $\lambda^*$  can be found by satisfying the complementary slackness

$$\lambda \left( \sum_i \mathbb{E}[\min(y_i^*(\lambda), D_{i,t})] - \frac{W}{T} \right) = 0.$$

As a result, in the case we only have an empirical demand distribution, the estimated  $\lambda^*$ , denoted by  $\hat{\lambda}_t$ , is found by solving the “empirical” complementary slackness

$$\lambda \left( \sum_i \hat{\mathbb{E}}[\min(\hat{y}_{i,t}(\lambda), \hat{D}_{i,t})] - \frac{W}{T} \right) = 0, \quad (\text{EC.5})$$

where  $\hat{D}_{i,t}$  is the demand with empirical distribution, and the notation  $\hat{\mathbb{E}}[\cdot]$  represents the expectation according to the empirical distribution. Since the left-hand side of (EC.5) is a left-continuous stepwise function, we let

$$\hat{\lambda}_t := \sup \left\{ \lambda \geq 0 : \sum_i \hat{\mathbb{E}}[\min(\hat{y}_{i,t}(\lambda), \hat{D}_{i,t})] - \frac{W}{T} \geq 0 \right\}. \quad (\text{EC.6})$$

The algorithm then simply applies the base-stock policy with level  $\hat{y}_{i,t}(\hat{\lambda}_t)$  (we write it as  $\hat{y}_{i,t}$  for brevity). The detailed algorithm is summarized in Algorithm 3.

---

**Algorithm 3:** EDAF(Empirical Distribution Algorithm for Full Demand)

---

**Input:** Starting warehouse capacity  $W_1 = W$ , initial store inventories  $I_{i,1}$ ,  $\forall i \in [N]$

**Output:** Policy  $\pi^{\text{EDAF}} = \{y_{i,t} = \max(\hat{y}_{i,t}, I_{i,t})\}$

```

1 for  $t \in [T]$  do
2   if  $t = 1$  then
3     | Set  $\hat{y}_{i,1} \in [\underline{D}, \overline{D}]$  arbitrarily for all  $i \in [N]$ 
4   else
5     | Compute  $\hat{\lambda}_t$  and  $\hat{y}_{i,t}$  as in (EC.4) and (EC.6)
6   end
7   if  $\sum_{i=1}^N (\hat{y}_{i,t} - I_{i,t})^+ \leq W_t$  then
8     | send  $(\hat{y}_{i,t} - I_{i,t})^+$  to each store  $i$ 
9     |  $W_{t+1} \leftarrow W_t - \sum_{i=1}^N (\hat{y}_{i,t} - I_{i,t})^+$ 
10    |  $I_{i,t+1} \leftarrow (\max(\hat{y}_{i,t}, I_{i,t}) - D_{i,t})^+$ 
11  else
12    | distribute  $W_t$  arbitrarily
13    |  $I_{i,t+1} \leftarrow (I_{i,t} - D_{i,t})^+$ 
14  end
15 end
```

---

To prove our claim in Proposition 1, we first show that the empirical estimates for order-up-to levels and the dual parameter  $\lambda$  do not deviate away from the real values with high probability. We then define  $V_{\text{unlim}}^\pi$  which is a hypothetical system of applying policy  $\pi$  with the same cost structure as the original OWMS problem, except without inventory constraint in the central warehouse. With the help of  $V_{\text{unlim}}^\pi$ , we decompose the regret into two parts: regret due to violating the capacity

constraint (i.e.,  $V^\pi - V_{\text{unlim}}^\pi$ ) and regret due to learning (i.e.,  $V_{\text{unlim}}^\pi - \tilde{V}^{\lambda^*}$ ). For the first part, we show that the regret that occurred because of violating the central warehouse capacity is at most  $\tilde{O}(N\sqrt{T})$ . Moreover, using the estimation properties, we show that regret due to learning is not larger than  $\tilde{O}(N\sqrt{T})$  as well.

We now give the detailed analysis of Proposition 1. Throughout our analysis, we will use  $\epsilon_t$  as an explicit variable. We define

$$\epsilon_t := \sqrt{\frac{\log(2/\delta_t)}{2t}},$$

for simplicity with respective probability values  $\delta_t$ . Before we go to the proof of Proposition 1, let us first prove the next lemma which reveals the consistency of the estimated parameters and costs.

LEMMA EC.5. *For any  $t \in [T]$ , estimates of the EDAF algorithm have the following properties with probability at least  $1 - O(N^{-2}T^{-2})$ .*

- i.  $|\lambda^* - \hat{\lambda}_{t+1}| \leq O(\epsilon_t)$
- ii.  $|\hat{y}_{i,t+1}(\hat{\lambda}_{t+1}) - y_i^*(\lambda^*)| \leq O(\epsilon_t)$
- iii.  $|C_i(\hat{y}_{i,t+1}(\hat{\lambda}_{t+1}); 0) - C_i(y_i^*(\lambda^*); 0)| \leq O(\epsilon_t)$

*Proof of Lemma EC.5* We first show that given  $|\hat{F}_{t+1}^i(d) - F^i(d)| < \epsilon_t$ , for all  $d$

$$|\hat{y}_{i,t+1}(\lambda) - y_i^*(\lambda)| \leq \frac{\epsilon_t}{\underline{f}}, \quad (\text{EC.7})$$

for any  $\lambda$ . In the rest of the proof, everything will be conditioned on event  $|\hat{F}_{t+1}^i(d) - F^i(d)| < \epsilon_t$ , which holds with probability at least  $1 - O(N^{-2}T^{-2})$ .

Let  $q = \kappa(\lambda)$ . To simplify notation, we omit the store number  $i$ . Additionally, with the abuse of notation, we denote  $F^{-1}(q) = \inf\{y : F(y) \geq q\}$ . We want to show that

$$|\hat{F}_{t+1}^{-1}(q) - F^{-1}(q)| \leq \frac{\epsilon_t}{\underline{f}}. \quad (\text{EC.8})$$

If this claim holds, then (EC.7) follows immediately. To this end, let  $d = F^{-1}(q)$ ,  $\hat{d} = \hat{F}_{t+1}^{-1}(q)$ , and  $\hat{q} = F(\hat{d})$ . Then  $|\hat{F}_{t+1}^{-1}(q) - F^{-1}(q)| = |F^{-1}(\hat{q}) - F^{-1}(q)| \leq |\hat{q} - q|/\underline{f}$  where  $\underline{f}$  is the lower bound of probability density function from Assumption 1 and the inequality is due to inverse function theorem. Note that we have  $|F(\hat{d}) - \hat{F}_{t+1}(\hat{d})| \leq \epsilon_t$  by the given event.

By definition of  $\hat{F}_{t+1}^{-1}$ , we always have  $\hat{F}_{t+1}(\hat{d}) \geq q$ . In the case that  $\hat{F}_{t+1}(\hat{d}) = q$ , we immediately have  $|\hat{F}_{t+1}^{-1}(q) - F^{-1}(q)| \leq |\hat{q} - q|/\underline{f} = |F(\hat{d}) - \hat{F}_{t+1}(\hat{d})|/\underline{f} = O(\epsilon_t)$ . Thus let us focus on the case that  $\hat{F}_{t+1}(\hat{d}) > q$ . Then we have another two cases:  $q < \hat{q}$  and  $q \geq \hat{q}$ . In the first case, note that by definition,  $\hat{F}_{t+1}$  has a staircase form. Thus  $\hat{F}_{t+1}(\hat{d}) > q$  implies that  $\hat{F}_{t+1}(\hat{d}^-) \leq q < \hat{F}_{t+1}(\hat{d})$ ,

where  $\hat{d}^- \leq \hat{d}$  is infinitesimally close to  $\hat{d}$ . By the fact that  $0 \leq q - \hat{F}_{t+1}(\hat{d}^-) < \hat{q} - \hat{F}_{t+1}(\hat{d}^-) = F(\hat{d}^-) - \hat{F}_{t+1}(\hat{d}^-) \leq \epsilon_t$  where the equality is by continuity of  $F$ , we have

$$|\hat{F}_{t+1}^{-1}(q) - F^{-1}(q)| \leq \frac{\hat{q} - q}{|f|} \leq \frac{\hat{q} - \hat{F}_{t+1}(\hat{d}^-)}{|f|} \leq \frac{\epsilon_t}{|f|} = O(\epsilon_t).$$

In the second case that  $q \geq \hat{q}$ ,

$$|\hat{F}_{t+1}^{-1}(q) - F^{-1}(q)| \leq \frac{q - \hat{q}}{|f|} < \frac{\hat{F}_{t+1}(\hat{d}) - F(\hat{d})}{|f|} = O(\epsilon_t),$$

and we are done with the proof of (EC.8).

*Proof of Part i.* It can be seen that the dual program is concave in  $\lambda$ . The gradient of  $\tilde{V}^\lambda/T$  is given as

$$\sum_{i=1}^N \mathbb{E}[\min(y_i^*(\lambda), D_{i,1})] - \frac{W}{T}, \quad (\text{EC.9})$$

according to Miao et al. (2022).

For a given  $\lambda$ , according to (EC.7), at period  $t+1$ , we have

$$|\mathbb{E}[\min(\hat{y}_{i,t+1}(\lambda), \hat{D}_{i,1})] - \mathbb{E}[\min(y_i^*(\lambda), D_{i,1})]| \leq \frac{\epsilon_t}{|f|}.$$

We define  $g(\lambda) := \sum_{i=1}^N \mathbb{E}[\min(y_i^*(\lambda), D_{i,1})] - W/T$  and its empirical estimate at time  $t$  as  $\hat{g}_{t+1}(\lambda) := \sum_{i=1}^N \hat{\mathbb{E}}[\min(\hat{y}_{i,t+1}(\lambda), \hat{D}_{i,t+1})] - W/T$ . As a result, we immediately have

$$|g(\lambda) - \hat{g}_t(\lambda)| \leq \frac{N\epsilon_t}{|f|}. \quad (\text{EC.10})$$

Note that by concavity of  $\tilde{V}^\lambda$ , we have  $g(\lambda)$  is monotone strictly (by Assumption 1) decreasing for all  $\lambda \geq 0$  with  $|g'(\lambda)| = \Theta(N)$  (as by Assumption 2, we never have  $c_i + \lambda \geq b_i$  for any  $i$ ), and  $\hat{g}_t(\lambda)$  is a step-wise function.

There are two cases for  $\lambda^*$ .

**Case 1:**  $\lambda^* > 0$ . By complementary slackness condition in Lemma 10 in Miao et al. (2022),  $\lambda^* = g^{-1}(0)$ . Then we just need to bound  $|\lambda^* - \hat{\lambda}_{t+1}| = |g^{-1}(0) - \hat{g}_{t+1}^{-1}(0)|$ . We just repeat the same proof for (EC.8) except we replace  $F, \hat{F}_{t+1}, \epsilon_t$  by  $g, \hat{g}_{t+1}, N\epsilon_t$  and we are done.

**Case 2:**  $\lambda^* = 0$ . In this case, we have  $g(0) \leq 0$  according to complementary slackness condition in Lemma 10 in Miao et al. (2022). However, we can always continuously extend  $g(\lambda)$  to  $\lambda < 0$  by, e.g., linear extension such that  $g(\lambda) = 0$  for some  $\lambda < 0$ . Similarly, we can do the same extension to  $\hat{g}_{t+1}(\lambda)$  such that  $|g(\lambda) - \hat{g}_{t+1}(\lambda)| \leq \frac{N\epsilon_t}{|f|}$  for any  $\lambda \in \mathbb{R}$ . Then we just repeat the same proof for (EC.8) except we replace  $F, \hat{F}_{t+1}, \epsilon_t$  by  $g, \hat{g}_{t+1}, N\epsilon_t$  (as  $|\lambda^* - \hat{\lambda}_{t+1}| \leq |g^{-1}(0) - \hat{g}^{-1}(0)|$ ) and we are done.

□

*Proof of Part ii.* When  $b_i - c_i - \lambda < 0$ , the order-up-to levels are zero; therefore, the claim trivially holds. Assume otherwise; the expression can be written as

$$|\hat{y}_{i,t+1}(\hat{\lambda}_{t+1}) - y_i^*(\lambda^*)| = \underbrace{|\hat{y}_{i,t}(\hat{\lambda}_{t+1}) - y_i^*(\hat{\lambda}_{t+1})|}_{(I)} + \underbrace{|y_i^*(\hat{\lambda}_{t+1}) - y_i^*(\lambda^*)|}_{(II)}.$$

Element (I) has  $O(\epsilon_t)$  upper bound from Line EC.7.

Element (II) can be written as

$$\begin{aligned} |F^{-1}(\kappa(\hat{\lambda}_t)) - F^{-1}(\kappa(\lambda^*))| &\leq \frac{1}{f(F^{-1}(\tilde{q}))} |\kappa(\hat{\lambda}_t) - \kappa(\lambda^*)| \\ &\leq \frac{\epsilon_t}{|f|}, \end{aligned}$$

where the first inequality follows from Taylor's Theorem for a  $\tilde{q}$  between  $\kappa(\hat{\lambda}_t)$  and  $\kappa(\lambda)$  and the inverse function theorem, and the second line is a result of *Part i.* due to linearity.

□

*Proof of Part iii.* We have

$$\begin{aligned} |C_i(\hat{y}_{i,t+1}(\hat{\lambda}_{t+1}); 0) - C_i(y_i^*(\lambda^*); 0)| &\leq L|\hat{y}_{i,t+1}(\hat{\lambda}_{t+1}) - y_i^*(\lambda^*)| \\ &\leq \frac{L\epsilon_t}{|f|}, \end{aligned}$$

where the first inequality is from Lipschitz continuity of  $C_i(y_i^*(\lambda); \lambda)$  and the second inequality is from part (ii).

□

The following proposition gives the performance analysis of the EDAF algorithm.

*Proof of Proposition 1* In this proof, we shall condition on the intersection of all events in Lemma EC.5, which happens with probability at least  $1 - O(N^{-1}T^{-1})$  by union bound. As a result, the regret on the complement of this event is at most  $O(1)$ , which can be neglected.

As we discussed, we have the regret can be bounded using benchmark  $\tilde{V}^{\lambda^*}$  as

$$\mathcal{R}(\pi) \leq V^\pi - \tilde{V}^{\lambda^*} + O(\sqrt{NT} + N).$$

As a result, it suffices to bound  $V^\pi - \tilde{V}^{\lambda^*}$ . To this end, we define  $V_{\text{unlim}}^\pi$  as a hypothetical system of applying policy  $\pi$  with the same cost structure as the original OWMS problem, except without inventory constraint in the central warehouse. Then we have

$$V^\pi - \tilde{V}^{\lambda^*} = V^\pi - V_{\text{unlim}}^\pi + V_{\text{unlim}}^\pi - \tilde{V}^{\lambda^*}.$$

As a result, it suffices to bound  $V^\pi - V_{\text{unlim}}^\pi$  and  $V_{\text{unlim}}^\pi - \tilde{V}^{\lambda^*}$  separately.

**Bound of  $V^\pi - V_{\text{unlim}}^\pi$ .** We first prove the following claim:

$$\begin{aligned} & \mathbb{E} \left[ \left( \sum_{i=1}^N \sum_{t=1}^T (\hat{y}_{i,t}(\hat{\lambda}_t) - I_{i,t})^+ - \sum_{t=1}^T \sum_{i=1}^N \min(y_i^*(\lambda^*), D_{i,t}) \right)^+ \right] \\ & \leq \tilde{O} \left( N \sum_{t=2}^T (\epsilon_t + \epsilon_{t-1}) + N \right). \end{aligned} \quad (\text{EC.11})$$

This holds because we have

$$\begin{aligned} (\hat{y}_{i,t+1}(\hat{\lambda}_{t+1}) - I_{i,t+1})^+ &= (\hat{y}_{i,t+1}(\hat{\lambda}_{t+1}) - (\max(\hat{y}_{i,t}(\hat{\lambda}_t), I_{i,t}) - D_{i,t})^+)^+ \\ &\leq \max(\hat{y}_{i,t+1}(\hat{\lambda}_{t+1}), (\hat{y}_{i,t+1}(\hat{\lambda}_{t+1}) - \max(\hat{y}_{i,t}(\hat{\lambda}_t), I_{i,t}) + D_{i,t})^+) \\ &\leq \max(y_i^*(\lambda^*) + \frac{\epsilon_t}{|f|}, \frac{\epsilon_t}{|f|} + \frac{\epsilon_{t-1}}{|f|} + D_{i,t}) \\ &\leq \max(y_i^*(\lambda^*), D_{i,1}) + \frac{\epsilon_t}{|f|} + \frac{\epsilon_{t-1}}{|f|}, \end{aligned}$$

where the second inequality is from Lemma EC.5.Part ii.

Now we are ready to bound  $V^\pi - V_{\text{unlim}}^\pi$ . Similar to Miao et al. (2022), the only possible extra cost in the real system compared with the unbounded system is the extra lost sales cost due to stockout in the central warehouse. Therefore, we have

$$\begin{aligned} V^\pi - V_{\text{unlim}}^\pi &\leq \overbrace{\bar{b} \mathbb{E} \left[ \left( \sum_{i=1}^N \sum_{t=1}^T (\hat{y}_{i,t}(\hat{\lambda}_t) - I_{i,t})^+ - W \right)^+ \right]}^{\text{penalty for central warehouse stockout}} \\ &\leq \bar{b} \mathbb{E} \left[ \left( \sum_{i=1}^N \sum_{t=1}^T (\hat{y}_{i,t}(\hat{\lambda}_t) - I_{i,t})^+ - \sum_{t=1}^T \sum_{i=1}^N \mathbb{E}[\min(y_i^*(\lambda^*), D_{i,t})] \right)^+ \right] + O(N) \\ &\leq \bar{b} \mathbb{E} \left[ \left( \sum_{i=1}^N \sum_{t=1}^T \min(y_i^*(\lambda^*), D_{i,t}) - \sum_{t=1}^T \sum_{i=1}^N \mathbb{E}[\min(y_i^*(\lambda^*), D_{i,t})] \right)^+ \right] \\ &\quad + \tilde{O} \left( N \sum_{t=2}^T (\epsilon_t + \epsilon_{t-1}) + N \right) + O(N) \\ &\leq O(\sqrt{NT}) + \tilde{O} \left( N \sum_{t=2}^T (\epsilon_t + \epsilon_{t-1}) + N \right) + O(N) \\ &\leq \tilde{O}(N\sqrt{T}), \end{aligned} \quad (\text{EC.12})$$

where the second inequality is from Lemma 10 in Miao et al. (2022), the third inequality is from (EC.11), the fourth inequality is from the upper bound by the variance of the sum of independent random variables, and the last inequality is by definition of  $\epsilon_t$ .

**Bound of  $V_{\text{unlim}}^\pi - \tilde{V}^{\lambda^*}$ .** But before we go to the specific bound for each part, let us prove an important claim: for any  $t \geq 2t_0 = 4\bar{D}^2 \log(NT)/\underline{D}^2$  and  $i \in [N]$ , with probability at least  $1 - O(N^{-2}T^{-2})$ ,

$$\max(\hat{y}_{i,t+1}(\hat{\lambda}_{t+1}), I_{i,t+1}) - y_i^*(\lambda^*) \leq O(\epsilon_t). \quad (\text{EC.13})$$

Notice that by our condition on events in Lemma EC.5, we have  $|\hat{y}_{i,t+1}(\hat{\lambda}_{t+1}) - y_i^*(\lambda^*)| \leq O(\epsilon_t)$  for all  $t$ . Thus this inequality does not hold only if  $I_{i,t+1}$  is very large. However, we shall show that this is very unlikely to happen. To this end, by Hoeffding's inequality (Hoeffding 1963)

$$\mathbb{P}\left(\sum_{s=t-t_0+2}^{t+1} D_{i,s} \geq t_0 \mathbb{E}[D_{i,1}] - \varepsilon\right) \geq 1 - \exp\left(-\frac{2\varepsilon^2}{t_0 \bar{D}}\right).$$

Let  $\varepsilon = \bar{D} t_0^{1/2} \log(NT)^{1/2}$ , then

$$\mathbb{P}\left(\sum_{s=t-t_1+2}^{t+1} D_{i,s} \geq t_0 \mathbb{E}[D_{i,1}] - \bar{D} t_0^{1/2} \log(NT)^{1/2}\right) \geq 1 - \frac{1}{N^2 T^2}. \quad (\text{EC.14})$$

Then on the event above,  $\sum_{s=t-t_0+2}^{t+1} D_{i,s} \geq t_0 \mathbb{E}[D_{i,1}] - \bar{D} t_0^{1/2} \log(NT)^{1/2} \geq \frac{1}{2} t_0 \mathbb{E}[D_{i,1}] \geq \bar{D}$  given that  $t_0 = 2\bar{D}^2 \log(NT)/\underline{D}^2$ . This means no matter what is the value of  $\max(\hat{y}_{i,t+1-t_0}(\hat{\lambda}_{t+1}), I_{i,t+1-t_0})$  (which is at most  $\bar{D}$ ), at some time  $s \in [t+1-t_0, t+1]$ , we will have  $I_{i,s} \leq \hat{y}_{i,s}(\hat{\lambda}_s)$ , which further implies that

$$\max(\hat{y}_{i,t+1}(\hat{\lambda}_{t+1}), I_{i,t+1}) \leq \max_{s \in [t+1-t_0, t+1]} \hat{y}_{i,s}(\hat{\lambda}_s) \leq y_i^*(\lambda^*) + O(\epsilon_{t+1-t_0}) \leq y_i^*(\lambda^*) + O(\sqrt{2}\epsilon_{t+1-t_0}),$$

given that  $t \geq 2t_0 = 4\bar{D}^2 \log(NT)/\underline{D}^2$ .

In the rest of the proof, we shall condition on event (EC.13) over all  $t \geq 4\bar{D}^2 \log(NT)/\underline{D}^2$  and  $i \in [N]$ , which happens with probability at least  $1 - O(N^{-1}T^{-1})$  by union bound (hence the regret on the complement is at most  $O(1)$  which can be neglected).

For the bound of  $V_{\text{unlim}}^\pi - \tilde{V}^{\lambda^*}$ , by definition of  $\tilde{V}^{\lambda^*}$  and complementary slackness condition in Lemma 10 in Miao et al. (2022), we have

$$\begin{aligned} V_{\text{unlim}}^\pi - \tilde{V}^{\lambda^*} &= \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}[C_i(\max(\hat{y}_{i,t}(\hat{\lambda}_t), I_{i,t}); 0) - C_i(y_i^*(\lambda^*); 0)] \\ &\quad + \sum_{i=1}^N c_i \mathbb{E}[(\max(\hat{y}_{i,T}(\hat{\lambda}_T), I_{i,T}) - D_{i,T})^+] \\ &\leq \tilde{O}\left(N \sum_{t=2t_0+1}^T \epsilon_t + 2Nt_0\right) + O(N) \\ &\leq \tilde{O}(N\sqrt{T}), \end{aligned} \quad (\text{EC.15})$$

where the first inequality is from (EC.13) and Lipschitz continuity of  $C_i(\cdot; 0)$ , and the last inequality is by definition of  $\epsilon_t, t_0$ .

Summarizing (EC.12) and (EC.15) gives us the final regret bound.

□

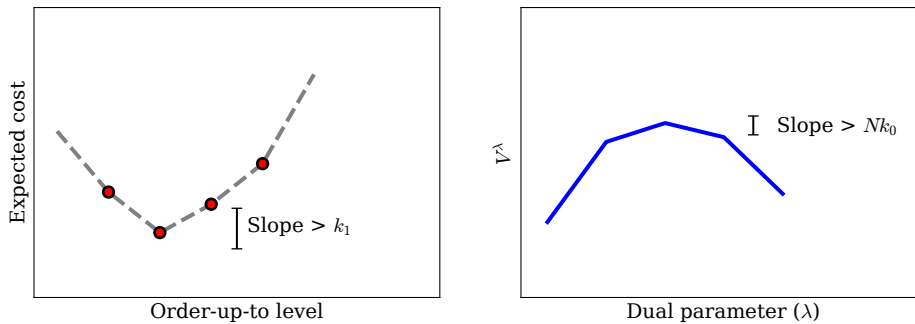
**A note on performance differences between DBS and EDAF algorithms:** Even though both DBS and EDAF algorithms result in  $\sqrt{T}$  growth with respect to time horizon, we would like to

note that observing uncensored demand leads to performance improvements compared to censored demand case. For instance, in the DBS algorithm, as the range of demand distribution (i.e.,  $\bar{D} - \underline{D}$ ) gets wider, the confidence band estimator that we generate to control the dual variable  $\lambda$  and the inventory-up-to-level decisions for each store becomes wider as well. This, in turn, results in longer search periods for the outer loop, which worsens the performance of the DBS algorithm. However, in the EDAF algorithm, we can estimate empirical demand distribution in an unbiased fashion, which helps us update both the dual variable and the inventory-up-to-level decisions for each store in each period without the need for a search period. This boosts the performance of EDAF relative to the performance of DBS significantly.

## EC.5. Technical Details for Discrete Demand

To analyze the discrete demand, let us denote several constants. Since we know that  $\tilde{V}^\lambda$  is piecewise linear in  $\lambda$  and  $C_i(y; \lambda^*)$  is piece-wise linear in  $y$ , let  $k_0 N, k_1 > 0$  be the minimum absolute value of nonzero slopes of  $\tilde{V}^\lambda$  and  $C_i(y; \lambda^*)$ , respectively. We give the formal assumption in Assumption EC.1. See Figure EC.1 for a graphical representation of  $k_0 N, k_1$ . The reason we have  $N$  with  $k_0$  is because changes of  $\lambda$  will change the value of the objective cost of all  $N$  stores. Note that although we assume the existence of such constants  $k_0, k_1$ , the implementation of our algorithm does not need to know the value of them.

**Assumption EC.1** For any  $\lambda \geq 0$ , and  $i \in [N]$ , there exist  $k_0, k_1 \in \mathbb{R}^+$  such that  $\min_{y_1, y_2 \in \mathbb{R}} \left| \frac{C_i(y_1; \lambda) - C_i(y_2; \lambda)}{y_1 - y_2} \right| \geq k_1$  and  $\min_{\lambda_1, \lambda_2 \in \mathbb{R}} \left| \frac{\tilde{V}^{\lambda_1} - \tilde{V}^{\lambda_2}}{\lambda_1 - \lambda_2} \right| \geq k_0 N$ .



**Figure EC.1** Minimum slopes of  $C_i(y; \lambda^*)$  and  $V^\lambda$ .

The general idea of the algorithm is as follows. Since the set of demand realizations is  $J$ , this means there exists some  $y_i(\lambda) \in J$  for any  $i \in [N], \lambda \geq 0$ . As a result, it suffices to first sample demand during  $C_0$  time periods (to be determined) for all stores by applying  $\bar{D}$  as the order-up-to level. The

purpose is that for any value of  $\lambda$  and  $j \in J$ ,  $\mathcal{C}_0$  samples of demand data can be used to simulate all candidate policies and get an estimate of  $\lambda$  sufficiently close to  $\lambda^*$  (denoted by  $\hat{\lambda}$ ). Likewise, the initial sampling can get rid of most of  $j \in J$ , which are *obviously suboptimal* for  $C_i(y; \hat{\lambda})$ . Then, the rest of the algorithm applies a binary search on the shrunk interval of  $y$  simultaneously over all  $i \in [N]$ . More specifically, denote the shrunk interval as  $[\hat{y}_i^1(\hat{\lambda}), \hat{y}_i^2(\hat{\lambda})]$ . We want to find the parameter  $\alpha \in [0, 1]$  so that  $\sum_i \mathbb{E}[\min\{D_{i,1}, \alpha \hat{y}_i^2(\hat{\lambda}) + (1 - \alpha) \hat{y}_i^1(\hat{\lambda})\}]$  is equal to  $W/T$  (i.e., the complementary slackness). The  $\hat{\lambda}$ ,  $[\hat{y}_i^1(\hat{\lambda}), \hat{y}_i^2(\hat{\lambda})]$  are computed as follows. First, denote  $\hat{C}_i(y; \lambda)$  as the empirically estimated  $C_i(y; \lambda)$  using the demand data in  $\mathcal{C}_0$  periods, and  $\hat{V}^\lambda = (w - \lambda)W + T \cdot \sum_i \min_y \hat{C}_i(y; \lambda)$ . Then  $\hat{\lambda} = \arg \max_\lambda \hat{V}^\lambda$  and  $\hat{y}_i^1(\hat{\lambda}) = \min\{j \in J : j \in \arg \min \hat{C}_i(j; \hat{\lambda})\}$ ,  $\hat{y}_i^2(\hat{\lambda}) = \max\{j \in J : j \in \arg \min \hat{C}_i(j; \hat{\lambda})\}$ . The whole algorithm is described in Algorithm 4.

---

**Algorithm 4:** Discrete Demand  $\alpha$ -Search (DD)

---

**Input:** All model parameters for the OWMS problem; parameters  $\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ ;

- 1 **Initialize:**  $l = 0, g_l = 0, L_l = 0, R_l = 1, \alpha_l = (L_l + R_l)/2$ ;
  - 2 **Initial sampling:** Set order-up-to levels to  $\bar{D}$  for  $\mathcal{C}_0$  periods for all stores;
  - 3 **Estimate:**  $\hat{\lambda}$ , and  $[\hat{y}_i^1(\hat{\lambda}), \hat{y}_i^2(\hat{\lambda})]$  by empirically minimizing  $\hat{V}^\lambda$  and  $\hat{C}_i(y; \lambda)$  for  $i \in [N]$ ;
  - 4 **while**  $T$  periods have not been reached and warehouse did not run out of stock **do**
  - 5     Set  $y_{i,l} = \hat{y}_i^1(\hat{\lambda})$  if  $[\hat{y}_i^1(\hat{\lambda}), \hat{y}_i^2(\hat{\lambda})]$  is a singleton
  - 6     else  $y_{i,l} = \hat{y}_i^1(\hat{\lambda}) + \alpha_l(\hat{y}_i^2(\hat{\lambda}) - \hat{y}_i^1(\hat{\lambda}))$  for all  $i \in [N]$ ;
  - 7     Set  $\nu_l = \mathcal{C}_1 \beta^l$ ;
  - 8     Apply  $y_{i,l}$  during  $\nu_l$  periods and observe  $z_{i,t} = \min(D_{i,t}, y_{i,l})$  for all  $i \in [N]$  ;
  - 9     Update  $g_l \leftarrow \sum_i \sum_{t \in \nu_l} z_{i,t} - W/T$  and  $[\underline{g}_l, \bar{g}_l] = g_l \pm \frac{\mathcal{C}_2}{\sqrt{\nu_l}}$ ;
  - 10    **if**  $\underline{g}_l \geq 0$  **then**
  - 11      |  $L_{l+1} \leftarrow \alpha_l, R_{l+1} \leftarrow R_l$ ;
  - 12    **else if**  $\bar{g}_l \leq 0$  **then**
  - 13      |  $L_{l+1} \leftarrow L_l, R_{l+1} \leftarrow \alpha_l$ ;
  - 14    **else**
  - 15      |  $L_{l+1} \leftarrow \text{Proj}_{[0,1]}(\alpha_l - \mathcal{C}_3/\sqrt{\nu_l}), R_{l+1} \leftarrow \text{Proj}_{[0,1]}(\alpha_l + \mathcal{C}_3/\sqrt{\nu_l})$ ;
  - 16    **end**
  - 17     $\alpha_{l+1} \leftarrow \frac{L_{l+1} + R_{l+1}}{2}$ ;
  - 18     $l \leftarrow l + 1$  ;
  - 19 **end**
- 

The performance of Algorithm 4 is summarized in the following theorem. For simplicity, we outline a sketch of its proof by ignoring similar techniques used in our main context.

THEOREM EC.1. Let  $\pi^{DD}$  be the policy applied by the Discrete Demand algorithm with,

$$\begin{aligned} \mathcal{C}_0 &= \sqrt{T}, \mathcal{C}_1 \geq 4/\beta^2, \mathcal{C}_2 \geq \sqrt{(3/2) \log(NT)}, \\ \mathcal{C}_3 &\geq 2\mathcal{C}_2 / \left( \sum_i (\hat{y}_i^2(\hat{\lambda}) - \hat{y}_i^1(\hat{\lambda}))(1 - F^i(\hat{y}_i^2(\hat{\lambda}))) \right), \beta \in (1, 4], \end{aligned}$$

If Assumption EC.1 is satisfied, then it has

$$\mathcal{R}(\pi^{DD}) \leq O(N\sqrt{T}).$$

*Sketch of Proof of Theorem EC.1.* We know that  $C_i(y; \lambda^*)$  is a piece-wise linear convex function with break-points at  $j \in J$ . It may have unique or multiple optima as a line segment. Moreover,  $\tilde{V}^\lambda$  is a piece-wise linear concave function with break-points as well. Without loss of generality, we assume that  $\lambda^*$  is a unique optimum. If  $\lambda^*$  is not unique, any  $\lambda$  from the optimal set will imply the same policy; thus, our results will be applicable to that case as well. Since the minimum absolute non-zero slope of the store cost function  $C_i(\cdot; \lambda^*)$  is  $k_1$  and the minimum absolute non-zero slope of  $V^\lambda$  is  $k_0N$ , if the initial sampling results in a resolution that is able to distinguish differences less than  $k_1$  and  $k_0N$  respectively, the subsequent binary search can easily be applied on optimal order-up-to levels minimizing  $C_i(\cdot; \lambda^*)$ . Therefore, we start by showing the ability to find a  $\hat{\lambda}$  that is close  $\lambda^*$  after the initial sampling phase and given  $\hat{\lambda}$ , the ability to find the optimal segment (or singleton)  $[y_i^1, y_i^2]$  (for convenience of notation, we use  $y_i^1, y_i^2$  to denote  $\hat{y}_i^1(\lambda^*), \hat{y}_i^2(\lambda^*)$ ) for each store  $i \in [N]$ . We note that during the exploration phase, the maximum order-up-to level  $\bar{D}$  is applied; thus, uncensored demand data is observed. Thanks to the discrete structure, the uncensored demand data gathered during this phase can be used to simulate all candidate policies  $j \in J$ .

Let  $\hat{C}_i(y; \lambda)$  and  $\hat{V}^\lambda$  be empirical expected costs using data in the initial sampling. We define the events for the empirical expected costs has an error of no more than  $\epsilon_1$  and  $\epsilon_2$  (to be specified) as

$$\mathcal{E}_1 := \{|C_i(y; \lambda) - \hat{C}_i(y; \lambda)| \leq \epsilon_1, \forall i \in [N], y \in J, \lambda \geq 0\},$$

and

$$\mathcal{E}_2 := \{|\tilde{V}^\lambda - \hat{V}^\lambda| \leq \epsilon_2, \forall \lambda \geq 0\}.$$

Note that by Hoeffding's inequality, it is straightforward to show that  $\mathcal{E}_1$  can hold with probability at least  $1 - O(T^{-1})$  with  $\epsilon_1 = \sqrt{\log(NT)/\mathcal{C}_0}$ . Similarly, we also have with probability at least  $1 - O(T^{-1})$ ,  $\mathcal{E}_2$  holds with  $\epsilon_2 = N\sqrt{\log(NT)/\mathcal{C}_0}$ .

On event  $\mathcal{E}_2$ , let  $\hat{\lambda}^*$  be a maximizer of  $\hat{V}^\lambda$ . It is not difficult to see that

$$\begin{aligned} & \tilde{V}^{\lambda^*} - \hat{V}^{\lambda^*} \\ &= \tilde{V}^{\lambda^*} - \hat{V}^{\lambda^*} + \hat{V}^{\lambda^*} - \hat{V}^{\hat{\lambda}^*} + \hat{V}^{\hat{\lambda}^*} - \tilde{V}^{\hat{\lambda}^*} \\ &\leq 2\epsilon_2. \end{aligned}$$

Thus, by taking  $\mathcal{C}_0 = \sqrt{T}$  and as  $T$  grows, we have  $2\epsilon_2 = o(Nk_0)$ , meaning that by our minimum non-zero slope assumption of  $\tilde{V}^\lambda$ , we have  $|\hat{\lambda}^* - \lambda^*| \leq 2\epsilon_2/(k_0N) = 2\sqrt{\log(NT)/\mathcal{C}_0}/k_0$ . Therefore, let  $\hat{y}_i^*, y_i^* \in J$  be arbitrary minimizer of  $C_i(\cdot; \hat{\lambda}^*), C_i(\cdot; \lambda^*)$  respectively, we have

$$\begin{aligned} & C_i(\hat{y}_i^*; \lambda^*) - C_i(y_i^*; \lambda^*) \\ &= C_i(\hat{y}_i^*; \lambda^*) - C_i(\hat{y}_i^*; \hat{\lambda}^*) + C_i(\hat{y}_i^*; \hat{\lambda}^*) - \hat{C}_i(\hat{y}_i^*; \hat{\lambda}^*) + \hat{C}_i(\hat{y}_i^*; \hat{\lambda}^*) - \hat{C}_i(y_i^*; \hat{\lambda}^*) \\ &\quad + \hat{C}_i(y_i^*; \hat{\lambda}^*) - C_i(y_i^*; \hat{\lambda}^*) + C_i(y_i^*; \hat{\lambda}^*) - C_i(y_i^*; \lambda^*) \\ &\leq O(4\sqrt{\log(NT)/\mathcal{C}_0}/k_0 + 2\sqrt{\log(NT)/\mathcal{C}_0}), \end{aligned}$$

where the last inequality is from  $|\hat{\lambda}^* - \lambda^*| \leq 2\epsilon_2/(k_0N) = 2\sqrt{\log(NT)/\mathcal{C}_0}/k_0$  and event  $\mathcal{E}_1$ . Therefore, taking  $\mathcal{C}_0 = \sqrt{T}$  with sufficiently large  $T$ , by the minimum slope assumption of  $C_i(\cdot; \lambda^*)$ , Algorithm 4 is able to identify the optimal  $[y_i^1, y_i^2]$  exactly for each  $C_i(\cdot; \lambda^*)$  on event  $\mathcal{E}_1, \mathcal{E}_2$ .

Now, we are ready to show the convergence of the binary search for  $\alpha$ .  $\alpha$ -Search tries to find the zero point of  $\mathcal{L}(\alpha) := \sum_i \mathbb{E}[\min\{D_{i,1}, y_i^1 + \alpha(y_i^2 - y_i^1)\}] - W/T$  over  $\alpha$  and it is a strictly increasing function with derivative  $\sum_i (y_i^2 - y_i^1)(1 - F^i(y_i^1 + \alpha(y_i^2 - y_i^1)))$ . Our analysis resembles the proof in the continuous case. To show pruning consistency, let us define

$$\mathcal{E}_3 := \{\mathcal{L}(\alpha) \in [\underline{g}_l, \bar{g}_l] \text{ for all } l\},$$

which is the event that the gradient is included in the interval  $[\underline{g}_l, \bar{g}_l]$  for all  $l$ . We shall show this event holds with probability at least  $1 - O(T^{-2})$ , and the rest of the proof shall be on this event.

$\hat{g}$  is an unbiased estimation of  $\mathcal{L}(\alpha)$ , and we have by Hoeffding's inequality (Hoeffding 1963)

$$\mathbb{P}(\mathcal{L}(\alpha) \notin [\hat{g} - \zeta, \hat{g} + \zeta]) \leq 2 \exp(-2n_l \zeta^2),$$

for any  $\zeta > 0$ . Setting  $\zeta = \sqrt{(3/2)\log(NT)}/\sqrt{n_l}$ , and having the constant  $\mathcal{C}_2 \geq \sqrt{(3/2)\log(NT)}$  we have  $1 - O(N^{-3}T^{-3})$  probability for the gradient included in the confidence interval. Applying union bound for  $T$  and  $N$ , we have

$$P(\mathcal{E}_3) = P(\mathcal{L}(\alpha) \in [\underline{g}_l, \bar{g}_l] \forall l) \geq 1 - O(N^{-2}T^{-2}).$$

Therefore, when  $\underline{g}_l > 0$ , then  $\mathcal{L}(\alpha) > 0$  and when  $\bar{g}_l < 0$ ,  $\mathcal{L}(\alpha) < 0$ .

Now, we claim that

$$R_l - L_l \leq \beta^{3/2} \mathcal{C}_3 \beta^{-l/2}, \quad (\text{EC.16})$$

and  $\alpha^* \in [L_l, U_l]$  for all  $l$ . We prove this claim as follows.

For  $l=0$ , we have  $\beta^{3/2} \mathcal{C}_3 \geq 1$ , thus, our claim holds. Assume that it holds for  $l$ , for  $l=l+1$ , we have one of the following conditions:

- if  $\underline{g} \geq 0$ , we have

$$R_{l+1} - L_{l+1} = (R_l - L_l)/2 \leq \beta^{3/2} \mathcal{C}_3 \beta^{-l/2} / 2 \leq \beta^{3/2} \mathcal{C}_3 \beta^{-(l+1)/2}, \quad (\text{EC.17})$$

where the first inequality is from the inductive hypothesis, and the last inequality is because  $\beta \leq 4$ . Also, since  $\alpha^* \in [L_l, R_l]$ ,  $\underline{g} \geq 0$  implies that  $\alpha^* \in [L_{l+1}, R_{l+1}]$ ;

- if  $\bar{g} \leq 0$ , Line EC.17 still holds. Moreover, this time  $\bar{g} \leq 0$  implies that  $\alpha^* \in [L_{l+1}, R_{l+1}]$ ;
- otherwise, note that  $0 \in [\underline{g}, \bar{g}]$  and  $\mathcal{L}(\alpha) \in [\underline{g}, \bar{g}]$ . Thus,  $|\mathcal{L}(\alpha)| \leq \frac{2\mathcal{C}_2}{\sqrt{\nu_l}}$ . Also, we have

$$\mathcal{L}'(\alpha) = \sum_i (y_i^2 - y_i^1) (1 - F^i(y_i^1 + \alpha(y_i^2 - y_i^1))) \geq \sum_i (y_i^2 - y_i^1) (1 - F^i(y_i^2)).$$

Then, by Taylor's theorem

$$\mathcal{L}(\alpha_l) = \mathcal{L}(\alpha_l) - \mathcal{L}(\alpha^*) = \mathcal{L}'(\bar{\alpha})(\alpha_l - \alpha^*),$$

where  $\bar{\alpha}$  is between  $\alpha_l$  and  $\alpha^*$ . Hence, we have

$$\begin{aligned} |\alpha_l - \alpha^*| &\leq \frac{|\mathcal{L}(\alpha)|}{\sum_i (y_i^2 - y_i^1) (1 - F^i(y_i^2))} \\ &\leq \frac{2\mathcal{C}_2 / \sqrt{\nu_l}}{\sum_i (y_i^2 - y_i^1) (1 - F^i(y_i^2))} \\ &\leq \mathcal{C}_3 / \sqrt{\nu_l}, \end{aligned}$$

where the last inequality is because  $\mathcal{C}_3 \geq 2\mathcal{C}_2 / (\sum_i (y_i^2 - y_i^1) F^i(y_i^1))$ . Therefore,  $\alpha^*$  is included in  $[L_{l+1}, R_{l+1}]$ . Also,  $R_{l+1} - L_{l+1} \leq 2\mathcal{C}_3 \beta^{-l/2} / \sqrt{\mathcal{C}_1} \leq \beta^{3/2} \mathcal{C}_3 \beta^{-(l+1)/2}$ , because  $\mathcal{C}_1 \geq 4/\beta^2$ . Hence, our claim holds.

In what follows, we will show the upper bound of the overall regret. We can decompose the regret as

$$\mathcal{R}(\pi^{DD}) \leq V^{DD} - V_{unlim}^{DD} + V_{unlim}^{DD} - \tilde{V}^{\lambda^*}, \quad (\text{EC.18})$$

where  $V_{unlim}^{DD}$  is the cost of applying Algorithm 4 in a hypothetical system with the same cost structure as the original OWMS problem, except without inventory constraint in the central warehouse. Therefore, our regret will be bounded by  $V^{DD} - V_{unlim}^{DD}$  and  $V_{unlim}^{DD} - \tilde{V}^{\lambda^*}$  respectively.

The regret incurred during the initial sampling phase is at most  $O(N\sqrt{T})$ . Therefore, our interest will be in the period after the initial sampling. Denote the first period of Discrete Demand  $\alpha$ -Search as  $t_0$ . Conditioned on  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , the estimated policy after initial sampling has the optimal cost. Hence,  $V_{unlim}^{DD} - \tilde{V}^{\lambda^*}$  is expected to be zero with high probability. On the other hand,  $\alpha$ -Search section works towards finding the optimal ordering amount. Therefore, we will show the upper bound of  $V^{DD} - V_{unlim}^{DD}$ , which is the regret sourced by warehouse constraint.

**Bound of  $V^{DD} - V_{unlim}^{DD}$ .** The discrepancy between  $V^{DD} - V_{unlim}^{DD}$  comes only from the possible lost sales due to stockout in the central warehouse. We have

$$\begin{aligned} V^{DD} - V_{unlim}^{DD} &\leq \bar{b} \mathbb{E} \left[ \left( \sum_{i=1}^N \sum_{t=t_0}^T (y_{i,t} - I_{i,t})^+ - W \right)^+ \right] \\ &\leq \bar{b} \mathbb{E} \left[ \left| \sum_{i=1}^N \sum_{t=t_0}^T (y_{i,t} - I_{i,t})^+ - \sum_{i=1}^N T_0 \cdot \mathbb{E}[\min\{y_i^*(\lambda^*), D_{i,1}\}] \right| \right] + O(N), \end{aligned} \quad (\text{EC.19})$$

where the second inequality is from Lemma 10 in Miao et al. (2022),  $y_{i,t}$  denotes the actual base-stock level from Algorithm DD, and  $T_0 := T - t_0$ . Then,

$$\begin{aligned} &\mathbb{E} \left[ \left| \sum_{i=1}^N \sum_{t=t_0}^T (y_{i,t} - I_{i,t})^+ - \sum_{i=1}^N T_0 \cdot \mathbb{E}[\min\{y_i^*(\lambda^*), D_{i,1}\}] \right| \right] \\ &= \mathbb{E} \left[ \left| \sum_{i=1}^N \sum_{t=t_0}^T \min\{y_{i,t}, D_{i,1}\} - \sum_{i=1}^N T_0 \cdot \mathbb{E}[\min\{y_i^*(\lambda^*), D_{i,1}\}] \right| \right] \\ &\leq \mathbb{E} \left[ \left| \sum_{i=1}^N \sum_{t=t_0}^T \min\{y_i^*(\lambda^*), D_{i,1}\} - \sum_{i=1}^N T_0 \cdot \mathbb{E}[\min\{y_i^*(\lambda^*), D_{i,1}\}] \right| \right] + \sum_{i=1}^N \sum_{l=0}^{\bar{l}} ((y_i^2 - y_i^1) \beta^{3/2} \mathcal{C}_3 \beta^{-l/2}) \\ &\leq O(\sqrt{NT}) + \sum_{i=1}^N \sum_{l=0}^{\bar{l}} ((y_i^2 - y_i^1) \beta^{3/2} \mathcal{C}_3 \beta^{-l/2}) \\ &\leq O(\sqrt{NT}) + \sum_{l=0}^{\bar{l}} O(N\beta^{-l/2}) \\ &\leq O(N\sqrt{T}), \end{aligned}$$

where the first inequality comes from EC.16, and the second inequality comes from the upper bound by the variance of the sum of independent random variables.  $\square$

## EC.6. Technical Details for Unbounded Demand

*Proof of Corollary 1* We basically replace  $\bar{D}$ 's in the main steps of the proof of Theorem 1 with  $\bar{y}$ , replace Lemma EC.1 with Lemma EC.6 and replace Lemma EC.2 with Lemma EC.7.

$\square$

LEMMA EC.6. *Given a fixed dual variable  $\lambda$ , and the constant  $\mathcal{C}_1 \geq \sqrt{3/2}(h_i + b_i - c_i)\sqrt{\log(T)}$ , with probability  $1 - O(T^{-2})$ , if Algorithm 1 runs during  $\nu$  periods for any store  $i \in [N]$  the following hold.*

- i.  $\sum_{t \in \mathcal{T}_\nu} (C_i(y_{i,t}; \lambda) - C_i(y_i^*(\lambda); \lambda)) \leq \log_2(\nu \bar{y}) (C_1^2 \underline{f}^{-1} h_i \log(\nu) + (h_i + b_i - c_i) + \bar{b} \bar{y})$  where  $\mathcal{T}_\nu$  is the set of time periods for Algorithm 1, and  $y_{i,t}$  is the actual base-stock level in time  $t$ .
- ii.  $|y_{i,l} - y_i^*(\lambda)| \leq \frac{1}{\nu} + \frac{c_1 \sqrt{\lceil \log_2(\nu \bar{y}) \rceil}}{h_i \underline{f} \sqrt{\nu - \lceil \log_2(\nu \bar{y}) \rceil}}$  where  $y_{i,l}$  is the final output of the base-stock level.
- iii.  $|\hat{d}_{i,l} - \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}]| \leq \left(\bar{y} + \frac{c_1}{h_i \underline{f}}\right) \sqrt{\frac{\log(NT) \lceil \log_2(\nu \bar{y}) \rceil}{\nu - \lceil \log_2(\nu \bar{y}) \rceil}} + \frac{1}{\nu}$  where  $\hat{d}_{i,l}$  is the final output of the estimated expected sales.
- iv.  $|\sum_{t \in \mathcal{T}_\nu} Z_{i,t} - \nu \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}]| \leq K_1 \log(NT) \sqrt{\nu} + K_2'$  where  $Z_{i,t}$  is the actual ordering in time  $t$  for store  $i$ ,  $K_1 = 1 + \sqrt{3}(h_i + b_i - c_i - \lambda)/(h_i \underline{f})$  and  $K_2' = 16\bar{y}^2 \log(NT)(\bar{l} + 1)/(\mathbb{E}[D_{i,1}])^2$ .

*Proof of Lemma EC.6* We note that similar to Lemma EC.1, the event  $\mathcal{E}_1$  holds with probability at least  $1 - O(T^{-2})$ , hence,  $y_i^*(\lambda)$  is always between  $L_l$  and  $R_l$ .

Also, we claim that the inventory overshooting can be consumed in  $O(\log(NT))$  periods. To show that, let the consumption duration of no ordering periods be  $t_0$ , by Hoeffding's inequality (Hoeffding 1963), we have

$$\mathbb{P}\left(\sum_{s=t-t_0+2}^{t+1} D_{i,s} \geq t_0 \mathbb{E}[D_{i,1}] - \epsilon\right) \geq 1 - \exp\left(-\frac{2\epsilon^2}{t_0 \sigma^2}\right).$$

Let  $\epsilon = \sigma \sqrt{t_0 \log(NT)}$ , then we can write,

$$\mathbb{P}\left(\sum_{s=t-t_0+2}^{t+1} D_{i,s} \geq t_0 \mathbb{E}[D_{i,1}] - \sigma \sqrt{t_0 \log(NT)}\right) \geq 1 - \frac{1}{N^2 T^2}.$$

Conditioned on the event above, we have  $\sum_{s=t-t_0+2}^{t+1} D_{i,s} \geq t_0 \mathbb{E}[D_{i,1}] - \sigma \sqrt{t_0 \log(NT)} \geq \frac{1}{2} t_0 \mathbb{E}[D_{i,1}] \geq \bar{y}$ , given that  $t_0 \geq \max\{4\sigma^2 \log(NT)/(\mathbb{E}[D_{i,1}])^2, 2\bar{y}/\mathbb{E}[D_{i,1}]\}$ . After applying union bound on  $N$  stores and  $T$  periods, our claim holds with probability  $1 - O(N^{-1}T^{-1})$ .

*Proof of Part i.* Let  $\mathcal{T}_l$  be the set of time periods during  $l$ . Suppose  $l < \lceil \log_2(\nu \bar{y}) \rceil$ , and consider  $l_0$ -th iteration in loop  $l$ . Given  $\mathcal{E}_1$ , we have that

$$|C'_i(y_{i,l}; \lambda)| \leq c_1 \frac{1}{\sqrt{l_0 - 1}}. \quad (\text{EC.20})$$

Recall that according to Assumption 1 the probability density has  $0 < \underline{f} \leq f^i(d) \leq \bar{f} < \infty$  for all  $i \in [N]$  and  $\underline{f}, \bar{f} \in \mathbb{R}$ . Note that  $C''_i(y; \lambda) \geq (h_i + b_i - c_i - \lambda) \underline{f} \geq h_i \underline{f}$  for any order-up-to level  $y \in [0, \bar{y}]$  and  $0 \leq \lambda < b_i - c_i$ . Thus, Taylor's theorem gives that

$$C_i(y_i^*(\lambda); \lambda) - C_i(y_{i,l}; \lambda) = C'_i(y_{i,l}; \lambda)(y_i^*(\lambda) - y_{i,l}) + C''_i(r; \lambda)(y_i^*(\lambda) - y_{i,l})^2 \leq 0,$$

where  $r \in \mathbb{R}$  is between  $y_i^*(\lambda)$  and  $y_{i,l}$ . Consequently,

$$C'_i(y_{i,l}; \lambda)(y_i^*(\lambda) - y_{i,l}) + h_i \underline{f} (y_i^*(\lambda) - y_{i,l})^2 \leq 0.$$

Rearranging the sides

$$h_i \underline{f} |y_i^*(\lambda) - y_{i,l}| \leq -C'_i(y_{i,l}; \lambda) \leq |C'_i(y_{i,l}; \lambda)| \leq C_1 \frac{1}{\sqrt{l_0 - 1}}.$$

Therefore we have

$$|y_i^*(\lambda) - y_{i,l}| \leq \frac{C_1}{h_i \underline{f} \sqrt{l_0 - 1}}. \quad (\text{EC.21})$$

For each loop  $l$ , let  $R_{\mathcal{T}_l} = \sum_{t \in \mathcal{T}_l} (C_i(y_{i,t}; \lambda) - C_i(y_i^*(\lambda); \lambda))$ .

**Case 1.**  $l < \lceil \log_2(\nu \bar{y}) \rceil$ : For any  $l_0$ -th time period in loop  $l$ , we have

$$\begin{aligned} C_i(y_{i,l}; \lambda) - C_i(y_i^*(\lambda); \lambda) &\leq C'_i(y_{i,l}; \lambda) |y_i^*(\lambda) - y_{i,l}| \leq C_1 \frac{1}{\sqrt{l_0 - 1}} \cdot \frac{C_1}{h_i \underline{f} \sqrt{l_0 - 1}} \\ &= \frac{C_1^2}{h_i \underline{f} (l_0 - 1)}, \end{aligned}$$

where the first inequality follows from convexity of  $C_i(\cdot; \lambda)$  and the second one follows from (EC.20) and (EC.21).

Summing the cycle regret gives

$$R_{\mathcal{T}_l} \leq \bar{b}\bar{y} + \sum_{l_0=2}^{|\mathcal{T}_l|} \frac{C_1^2}{h_i \underline{f} (l_0 - 1)} \leq C_1^2 \underline{f}^{-1} h_i^{-1} \log(\nu) + \bar{b}\bar{y}.$$

**Case 2**  $l = \lceil \log_2(\nu \bar{y}) \rceil$ : Note that  $R_l - L_l \leq \bar{y}(\frac{1}{2})^{\log_2(\nu \bar{y})} = \frac{1}{\nu}$ . Hence,  $|y_i^*(\lambda) - y_{i,l}| \leq 1/\nu$ . Then, by convexity,

$$C_i(y_{i,l}; \lambda) - C_i(y_i^*(\lambda); \lambda) \leq C'_i(y_{i,l}; \lambda) |y_i^*(\lambda) - y_{i,l}| \leq \frac{h_i + b_i - c_i}{\nu};$$

thus the summation over the cycle gives that

$$R_{\mathcal{T}_l} \leq \frac{h_i + b_i - c_i}{\nu} n_l \leq h_i + b_i - c_i.$$

Combining the two cases, we have

$$\begin{aligned} \sum_{t=1}^{\nu} (C_i(y_{i,t}; \lambda) - C_i(y_i^*(\lambda); \lambda)) &\leq \sum_l R_{\mathcal{T}_l} \\ &\leq \log_2(\nu \bar{y}) (C_1^2 \underline{f}^{-1} h_i \log(\nu) + (h_i + b_i - c_i) + \bar{b}\bar{y}), \end{aligned}$$

which gives the result.

*Proof of Part ii.* Let  $l^*$  be the cycle with largest  $n_l$ .

**Case 1.**  $l^* < \lceil \log_2(\nu \bar{y}) \rceil$ : Let  $\bar{l}$  be the latest cycle. We have that

$$n_{l^*} \geq \frac{\nu}{\bar{l}} \geq \frac{\nu}{\lceil \log_2(\nu \bar{y}) \rceil},$$

where the first inequality follows from the pigeonhole principle. Then, by inequality (EC.21), we have

$$|y_i^*(\lambda) - y_{i,l^*}| \leq \frac{\mathcal{C}_1}{h_{i,f}\sqrt{n_{l^*} - 1}} \leq \frac{\mathcal{C}_1 \sqrt{\lceil \log_2(\nu \bar{y}) \rceil}}{h_{i,f}\sqrt{\nu - \lceil \log_2(\nu \bar{y}) \rceil}}.$$

**Case 2.**  $l^* = \lceil \log_2(\nu \bar{y}) \rceil$ : Then we have by binary search,  $|y_i^*(\lambda) - y_{i,l^*}| \leq 1/\nu$ .

Combining the two cases, we have

$$|y_{i,l^*}(\lambda) - y_i^*(\lambda)| \leq \frac{1}{\nu} + \frac{\mathcal{C}_1 \sqrt{\lceil \log_2(\nu \bar{y}) \rceil}}{h_{i,f}\sqrt{\nu - \lceil \log_2(\nu \bar{y}) \rceil}}.$$

*Proof of Part iii.* Let loop  $l^*$  be the one with the largest  $n_l$ .  $\hat{d}_{i,l^*}$  is defined as

$$\hat{d}_{i,l^*} = \frac{\sum_{t \in \mathcal{T}_{l^*}} \min\{y_{i,l^*}, D_{i,t}\}}{n_{l^*}}.$$

According to Hoeffding's inequality, we have that with probability at least  $1 - O(N^{-1}T^{-2})$ ,

$$\left| \hat{d}_{i,l^*} - \mathbb{E}[\min\{y_{i,l^*}, D_{i,1}\}] \right| \leq \bar{y} \sqrt{\frac{\log(NT)}{n_{l^*}}}.$$

Then by part *ii.*, we have

$$\begin{aligned} & \left| \hat{d}_{i,l^*} - \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}] \right| \\ & \leq \left| \hat{d}_{i,l^*} - \mathbb{E}[\min\{y_{i,l^*}, D_{i,1}\}] \right| + \left| \mathbb{E}[\min\{y_{i,l^*}, D_{i,1}\}] - \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}] \right| \\ & \leq \left| \hat{d}_{i,l^*} - \mathbb{E}[\min\{y_{i,l^*}, D_{i,1}\}] \right| + |y_{i,l^*} - y_i^*(\lambda)| \\ & \leq \bar{y} \sqrt{\frac{\log(NT)}{n_{l^*}}} + \frac{1}{\nu} + \frac{\mathcal{C}_1 \sqrt{\lceil \log_2(\nu \bar{y}) \rceil}}{h_{i,f}\sqrt{\nu - \lceil \log_2(\nu \bar{y}) \rceil}}. \end{aligned}$$

Since  $n_{l^*}$  is the largest among all  $n_l$ , and  $l \leq \lceil \log_2(\nu \bar{y}) \rceil$ , we must have  $n_{l^*} \geq \nu / \lceil \log_2(\nu \bar{y}) \rceil$ , implying that

$$\left| \hat{d}_{i,l^*} - \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}] \right| \leq \left( \bar{y} + \frac{\mathcal{C}_1}{h_{i,f}} \right) \sqrt{\frac{\log(NT) \lceil \log_2(\nu \bar{y}) \rceil}{\nu - \lceil \log_2(\nu \bar{y}) \rceil}} + \frac{1}{\nu}.$$

*Proof for Part iv.*

We have

$$\begin{aligned} \sum_{t \in \mathcal{T}_\nu} Z_{i,t} &= \sum_l \sum_{t \in \mathcal{T}_l} Z_{i,t} \\ &\leq \sum_l \sum_{t \in \mathcal{T}_l} \min\{y_{i,l}, D_{i,t}\} + \bar{y}(\bar{l} + 1). \end{aligned}$$

where  $\bar{l}$  is the index of the last cycle, and the last inequality is due to possible extra ordering at the beginning of each cycle  $l$ . Similarly, we have

$$\sum_{t \in \mathcal{T}_\nu} Z_{i,t} \geq \sum_l \sum_{t \in \mathcal{T}_l} \min\{y_{i,l}, D_{i,t}\} - 16\bar{y}^2 \log(NT)(\bar{l} + 1) / (\mathbb{E}[D_{i,1}])^2,$$

where  $16\bar{y}^2 \log(NT)/(\mathbb{E}[D_{i,1}])^2$  is from possible no-order periods in each loop.

Then

$$\begin{aligned}
& \left| \sum_{t \in \mathcal{T}_\nu} Z_{i,t} - \nu \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}] \right| \\
& \leq \left| \sum_l \sum_{t \in \mathcal{T}_l} \min\{y_{i,l}, D_{i,t}\} - \sum_{l=0}^{\bar{l}} n_l \mathbb{E}[\min\{y_{i,l}, D_{i,1}\}] \right| \\
& \quad + \sum_{l=0}^{\bar{l}} n_l |\mathbb{E}[\min\{y_{i,l}, D_{i,1}\}] - \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}]| + 16\bar{y}^2 \log(NT)(\bar{l}+1)/(\mathbb{E}[D_{i,1}])^2 \\
& \leq \sum_{l=0}^{\bar{l}} \left( \sqrt{\log(NT)n_l} \right) + \sum_{l=0}^{\bar{l}} n_l |\mathbb{E}[\min\{y_{i,l}, D_{i,1}\}] - \mathbb{E}[\min\{y_i^*(\lambda), D_{i,1}\}]| + 16\bar{y}^2 \log(NT)(\bar{l}+1)/(\mathbb{E}[D_{i,1}])^2 \\
& \leq \sqrt{\log(NT)\nu} + \frac{\nu \mathcal{C}_1 \sqrt{\lceil \log_2(\nu \bar{D}) \rceil}}{h_i \underline{f} \sqrt{\nu - \lceil \log_2(\nu \bar{D}) \rceil}} + 16\bar{y}^2 \log(NT)(\bar{l}+1)/(\mathbb{E}[D_{i,1}])^2 \\
& \leq \sqrt{\log(NT)\nu} + \frac{\mathcal{C}_1 \sqrt{2 \lceil \log_2(\nu \bar{D}) \rceil}}{h_i \underline{f}} \sqrt{\nu} + 16\bar{y}^2 \log(NT)(\bar{l}+1)/(\mathbb{E}[D_{i,1}])^2 \\
& \leq K_1 \log(NT) \sqrt{\nu} + K_2,
\end{aligned}$$

where the second inequality is satisfied with probability at least  $1 - O(N^{-1}T^{-2})$  using Hoeffding's inequality, the third inequality holds according to a similar argument as in Part *ii* and the constants  $K_1 = 1 + \frac{\sqrt{3}(h_i + b_i - c_i - \lambda)}{h_i \underline{f}}$  and  $K_2 = 16\bar{y}^2 \log(NT)(\bar{l}+1)/(\mathbb{E}[D_{i,1}])^2$ .

□

LEMMA EC.7. *Let*

$$\begin{aligned}
\mathcal{C}_0 & \geq \max(4/\beta^2, 2 \lceil \log_2(T\bar{y}) \rceil) \\
\mathcal{C}_2 & \geq \left( \bar{y} + \frac{\mathcal{C}_1}{h_i \underline{f}} \right) \sqrt{2 \log(NT) \lceil \log_2(T\bar{y}) \rceil} \\
\mathcal{C}_3 & \geq \max \left( \sqrt{\mathcal{C}_0} \min_i (b_i - c_i), 2\mathcal{C}_2 / \min_i \frac{h_i^2}{(h_i + b_i - c_i)^3 \bar{f}} \right) \\
& \beta \in (1, 4],
\end{aligned}$$

then with probability at least  $1 - O(T^{-1})$ , we have

$$|\hat{\lambda}_\tau - \lambda^*| \leq \mathcal{C}_3 \beta^{-\tau/2+3/2},$$

for all  $\tau$ .

*Proof of Lemma EC.7* Define the event

$$\mathcal{E}_2 := \left\{ \left| \sum_{i \in [N]} \hat{d}_{i,\tau} - \sum_{i \in [N]} \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] \right| \leq \mathcal{C}_2 N / \sqrt{\nu_\tau}, \forall \tau \right\}.$$

Note that

$$\left| \sum_{i \in [N]} \hat{d}_{i;\tau} - \sum_{i \in [N]} \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] \right| \leq \sum_{i \in [N]} \left| \hat{d}_{i;\tau} - \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] \right| \leq \mathcal{C}_2 N / \sqrt{\nu_\tau},$$

where the last inequality holds with probability at least  $1 - O(T^{-1})$  according to Lemma EC.6.(iii.) by union bound over  $\tau$  and the fact that  $\nu_0 = \mathcal{C}_0 \geq 2 \lceil \log_2(T\bar{y}) \rceil$  and  $\mathcal{C}_2 \geq \left(\bar{y} + \frac{c_1}{h_i \bar{f}}\right) \sqrt{2 \log(NT) \lceil \log_2(T\bar{y}) \rceil}$ . Thus event  $\mathcal{E}_2$  holds with probability at least  $1 - O(T^{-1})$ . Our argument will be conditioned on event  $\mathcal{E}_2$  holds.

By a bit abuse of notation, let us write  $\tilde{V}^\lambda$  as its normalized version (by  $T$ , i.e.,  $\tilde{V}^\lambda/T$ ). For notation convenience, let  $\tilde{V}'(\lambda) := \sum_i \mathbb{E}[\min(y_i^*(\lambda), D_{i,1})] - W/T = \frac{d\tilde{V}^\lambda}{d\lambda}$ . With  $\mathcal{E}_2$  holds, this shows that  $\tilde{V}'(\hat{\lambda}_\tau) = \sum_{i \in [N]} \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] - W/T \in [\underline{v}_\tau, \bar{v}_\tau]$  for all  $\tau$ . Moreover, we have, by inverse function theorem and Assumption 1 and Assumption 2,

$$\begin{aligned} \frac{d^2 \tilde{V}^\lambda}{d\lambda^2} &= - \sum_{i \in [N]} \frac{h_i}{(h_i + b_i - c_i - \lambda)^2 f^i(y_i^*(\lambda))} (1 - F^i(y_i^*(\lambda))) \\ &= - \sum_{i \in [N]} \frac{h_i^2}{(h_i + b_i - c_i - \lambda)^3 f^i(y_i^*(\lambda))} \\ &\leq - \sum_{i \in [N]} \frac{h_i^2}{(h_i + b_i - c_i)^3 \bar{f}} \leq -N \min_{i \in [N]} \frac{h_i^2}{(h_i + b_i - c_i)^3 \bar{f}}. \end{aligned}$$

Let us extend  $\tilde{V}'(\lambda)$  to the domain of negative  $\lambda$ . The way we do it is by extending  $\tilde{V}'(\lambda)$  linearly with the rate of  $\frac{d^2 \tilde{V}^0}{d\lambda^2}$  so that  $\tilde{V}'(\lambda)$  is differentiable after extension and we still have the same upper bound of  $\frac{d^2 \tilde{V}^\lambda}{d\lambda^2}$  for all  $\lambda \leq \min_i(b_i - c_i)$ . Note that since  $\frac{d^2 \tilde{V}^0}{d\lambda^2} < 0$ , we have  $\tilde{V}'(\lambda) \rightarrow \infty$  if  $\lambda \rightarrow -\infty$ , which implies, together with Assumption 2, that there always exists  $\tilde{\lambda} \in \mathbb{R}$  such that  $\tilde{V}'(\tilde{\lambda}) = 0$ . Obviously, if  $\tilde{\lambda} < 0$ , we have  $\lambda^* = 0$  and otherwise  $\lambda^* = \tilde{\lambda}$ .

For the rest of the proof, we **claim** that  $U_\tau - L_\tau \leq \beta^{3/2} \mathcal{C}_3 \beta^{-\tau/2}$  and  $\lambda^* \in [L_\tau, U_\tau]$  for all  $\tau$ . Obviously, if this claim is true, we immediately have our result. We prove this by induction. This statement is obviously true for  $\tau = 0$  by assuming  $\beta^{3/2} \mathcal{C}_3 \geq \min_i(b_i - c_i)$ . For  $\tau = 1$ , we have one of the following conditions.

If line 7 happens, we have  $U_\tau, L_\tau$  unchanged. Then just take  $\beta^{3/2} \mathcal{C}_3 \beta^{-1/2} \geq \min_i(b_i - c_i)$  and we are good. If line 10 happens, we immediately have  $\lambda^* = 0$  and  $U_\tau = L_\tau = 0$ , so we are good. If line 13 happens, since  $|\hat{\lambda}_0 - \lambda^*| = |\lambda^*| \leq \min_i(b_i - c_i) \leq \mathcal{C}_3 / \sqrt{\mathcal{C}_0}$  where the last inequality is because  $\mathcal{C}_3 \geq \sqrt{\mathcal{C}_0} \min_i(b_i - c_i)$ . Thus taking  $U_1 = \text{Proj}_{[0, \min_i(b_i - c_i)]}(\hat{\lambda}_0 + \mathcal{C}_3 / \sqrt{\mathcal{C}_0})$  and  $L_1 = \text{Proj}_{[0, \min_i(b_i - c_i)]}(\hat{\lambda}_0 - \mathcal{C}_3 / \sqrt{\mathcal{C}_0})$  has  $\lambda^* \in [L_1, U_1]$ , and

$$U_1 - L_1 \leq 2\mathcal{C}_3 / \sqrt{\mathcal{C}_0} \leq \beta^{3/2} \mathcal{C}_3 \beta^{-1/2},$$

because  $\mathcal{C}_0 \geq 4/\beta^2$ , which again proves the claim.

For the inductive step, suppose it holds for  $\tau$ ; let us consider  $\tau + 1$ . The procedure is almost the same as for  $\tau = 1$ .

If condition line 7 or 10 is satisfied for  $\tau$ , then by algorithm design, we have

$$U_{\tau+1} - L_{\tau+1} = (U_\tau - L_\tau)/2 \leq \beta^{3/2} \mathcal{C}_3 \beta^{-\tau/2} / 2 \leq \beta^{3/2} \mathcal{C}_3 \beta^{-(\tau+1)/2},$$

where the first inequality is from the inductive hypothesis, and the last inequality is because  $\beta \leq 4$ .

Moreover, since  $\tilde{V}'(\hat{\lambda}_\tau) = \sum_{i \in [N]} \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}] - W/T \in [\underline{v}_\tau, \bar{v}_\tau]$ , line 7 (10) implies that  $\hat{\lambda}_\tau \leq \lambda^*$  ( $\hat{\lambda}_\tau \geq \lambda^*$ ), which shows that  $\lambda^* \in [L_{\tau+1}, U_{\tau+1}]$ . Thus our claim is proved.

If condition line 13 is satisfied for  $\tau$ , this implies that  $0, \tilde{V}'(\hat{\lambda}_\tau) \in [\underline{v}_\tau, \bar{v}_\tau]$  and thus  $|\tilde{V}'(\hat{\lambda}_\tau)| \leq 2\mathcal{C}_2 N / \sqrt{\nu_\tau}$ . Note that by Taylor's theorem on  $\tilde{V}'(\lambda)$ , we have

$$\tilde{V}'(\hat{\lambda}_\tau) = \tilde{V}'(\hat{\lambda}_\tau) - \tilde{V}'(\tilde{\lambda}) = \frac{d^2 \tilde{V}^{\tilde{\lambda}}}{d\lambda^2} (\hat{\lambda}_\tau - \tilde{\lambda}),$$

where  $\tilde{\lambda}$  is between  $\hat{\lambda}_\tau$  and  $\tilde{\lambda}$ . This implies that

$$\begin{aligned} |\hat{\lambda}_\tau - \lambda^*| &\leq |\hat{\lambda}_\tau - \tilde{\lambda}| \\ &\leq \frac{1}{N \min_i \frac{h_i^2}{(h_i + b_i - c_i)^{3\bar{f}}}} \left| \tilde{V}'(\hat{\lambda}_\tau) \right| \\ &\leq \frac{2\mathcal{C}_2 / \sqrt{\nu_\tau}}{\min_i \frac{h_i^2}{(h_i + b_i - c_i)^{3\bar{f}}}} \\ &\leq \mathcal{C}_3 / \sqrt{\nu_\tau}, \end{aligned}$$

where the last inequality is because  $\mathcal{C}_3 \geq 2\mathcal{C}_2 / \min_i \frac{h_i^2}{(h_i + b_i - c_i)^{3\bar{f}}}$ . Thus taking  $U_{\tau+1} = \text{Proj}_{[0, \min_i (b_i - c_i)]}(\hat{\lambda}_\tau + \mathcal{C}_3 \beta^{-\tau/2} / \sqrt{\mathcal{C}_0})$  and  $L_{\tau+1} = \text{Proj}_{[0, \min_i (b_i - c_i)]}(\hat{\lambda}_\tau - \mathcal{C}_3 \beta^{-\tau/2} / \sqrt{\mathcal{C}_0})$  has  $\lambda^* \in [L_{\tau+1}, U_{\tau+1}]$ , and

$$U_{\tau+1} - L_{\tau+1} \leq 2\mathcal{C}_3 \beta^{-\tau/2} / \sqrt{\mathcal{C}_0} \leq \beta^{3/2} \mathcal{C}_3 \beta^{-(\tau+1)/2},$$

because  $\mathcal{C}_0 \geq 4/\beta^2$ , which again proves the claim.

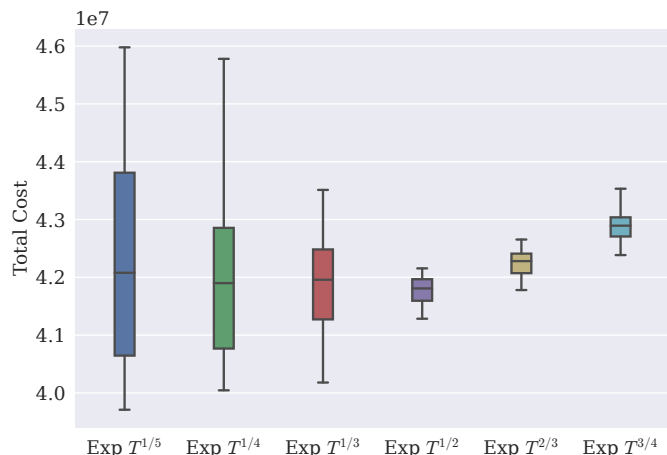
□

## EC.7. Additional Computational Experiments

### EC.7.1. Duration of Exploration Phase in Exp Algorithm

We conduct additional analyses by varying exploration duration in the explore-then-commit (Exp) algorithm. We achieve this by setting the exploration phase to  $T^\zeta$ , where  $\zeta$  is between 0 and 1, and then evaluate its performance for different  $\zeta$  values. In Figure EC.2, we show the mean performance of the Exp algorithm with respect to different  $\zeta$  values.

As shown in Figure EC.2, for a very small  $\zeta$  value, the exploration phase becomes very limited; hence, the explore-then-commit (Exp) algorithm does not explore enough to learn the true demand



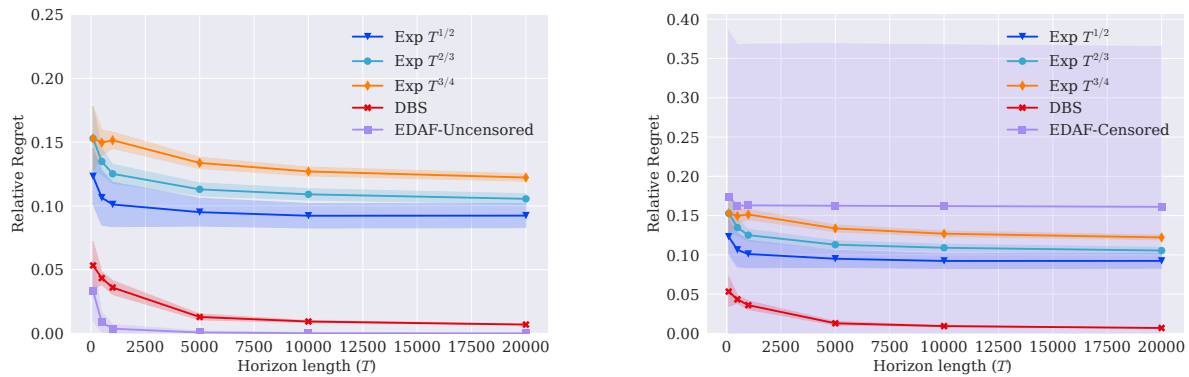
**Figure EC.2** The performance of Exp algorithms with respect to various  $\zeta$  values

distribution. For low  $\zeta$  values, this yields lower performance for the Exp algorithm in terms of both average and variance. As  $\zeta$  value increases, the Exp algorithm learns the true demand distribution more and more accurately. This decreases the variance and yields higher average performance for the Exp algorithm. However, for very high  $\zeta$  values, even though true demand parameters are estimated more accurately, which leads to reduced variance, the algorithm does not spend enough time to exploit the policy tuned for true demand parameters. As such, for the high value of  $\zeta$ , the average performance of the Exp algorithm degrades, whereas the variance of the Exp algorithm improves. To summarize, while the variance of the performance of the Exp algorithm decreases in  $\zeta$ , we observe U-shape behavior for the average performance of the Exp algorithm with respect to  $\zeta$ . In our numerical study, we consider the three  $\zeta$  values (i.e.,  $\frac{1}{2}$ ,  $\frac{2}{3}$ , and  $\frac{3}{4}$ ) that yield the lowest variance.

### EC.7.2. Performance of the Empirical Distribution Algorithm for Full Demand (EDAF)

In this experiment, we evaluate the impact of censored demand observations on the performance of our approach by comparing the EDAF algorithm developed for the uncensored demand case to the DBS algorithm developed for the censored case as well as other exploration-based heuristics (referred to as Exp  $T^\zeta$ ). Along these lines, we conducted two experiments. The left panel of Figure EC.3 shows the performance of all the algorithms under an uncensored data regime, i.e., assuming that we can observe demand observations even if we have run out of inventory, whereas the right panel shows the performance of the algorithms when we have censored demand observations. As shown in these figures, in terms of relative regret measure, the EDAF algorithm performs the best among all the algorithms under the uncensored demand case. This is quite intuitive, given the fact

that the EDAF algorithm is optimized for the benchmark case and does not invest in exploration. On the other hand, it performs the worst when the demand is censored. We particularly see that EDAF has a high bias and it suffers from variance severely. Using the EDAF algorithm under the censored demand case yields approximately a 15% increase in relative regret. Evidently, this strongly motivates the need to consider demand censoring in the algorithm design, which is exactly the main objective of our proposed DBS algorithm.



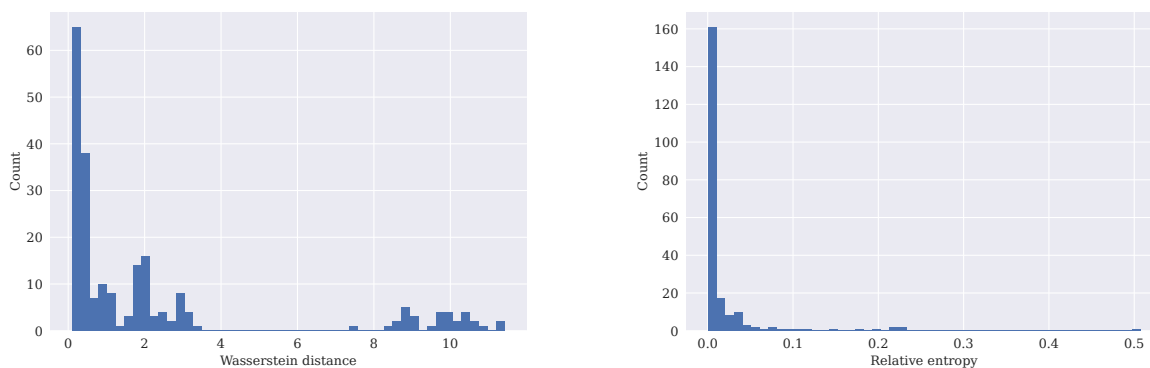
**Figure EC.3** Comparison between performances of all algorithms under the uncensored regime (left panel) and censored demand case (right panel).

### EC.7.3. Additional Information for Real-Data Experiments

The industry partner we collaborated with in this project is a fast-fashion company that operates more than 300 stores. The dataset we obtained from our industry partner consists of weekly sales and inventory data of more than 30,000 seasonal items that were sold in this network of stores between 2016 and 2019. Due to the extensive nature of this dataset, we select five representative stores in this network and focus only on the sales of 125 items randomly chosen from the women’s clothes category between July 2018 and June 2019. To assess the impact of different product characteristics on the performance of our approach, we divided 125 items into three subcategories. Namely, the items whose price is less than 25% percentile, between 25% and 75% percentile, and greater than 75% percentile are categorized as low-, medium-. and high-priced items, respectively. Similarly, the items whose sales is less than 25% percentile, between 25% and 75% percentile, and greater than 75% percentile are categorized as low-, medium-. and high-volume items.

For each item in our dataset, we first estimate the underlying demand distribution, then simulate DBS and Exp algorithms and calculate their regrets. In what follows, we provide the details of the estimation procedure. Since our data includes sales (i.e., censored demand observations), we

utilized two distinct estimation approaches to account for bias due to censored demand observations. In our first approach, we used a non-parametric method called "the reduced sample method" which was developed in Kaplan and Meier (1958). This method disregards censored samples and estimates distributions with the remaining sample. The estimations obtained from this method could be troublesome with frequent stock-outs because it does not make use of the entire dataset. Therefore, as a second approach, we used parametric methods developed by Nahmias (1994) and Conrad (1976). These methods apply adjustments to the reduced sample estimates of the parameters. Nahmias (1994) assumes Normal demand and Conrad (1976) assumes Poisson distributed demand. Last but not least, we also considered exponential, gamma, truncated normal, Weibull, and chi distributions, among other parametric distributions. The parameters of these distributions are found by maximum likelihood estimation based on the reduced sample method. The normal distribution largely fits the data quite well. To assess the goodness of fit, we compare the normal distributions generated by the two estimation procedures. Figure EC.4 shows the histogram of Wasserstein distances and relative entropies between Normal distributions computed under the reduced sample method and Normal distributions computed by the adjustment described in Nahmias (1994). As shown in Figure EC.4, one can observe that in the majority of instances, the two methods generate similar distributions.



**Figure EC.4** Histogram of Wasserstein distances (left panel) and relative entropies (right panel) between Normal distributions estimated based on the reduced sample and Normal distributions estimated based on the adjustment procedure described in Nahmias (1994).

Finally, we tested the validity of candidate distributions (except the Poisson distribution) using the Kolmogorov–Smirnov test. Kolmogorov–Smirnov test compares a sample with a distribution. The null hypothesis is

$$H_0 : \text{the sample comes from the reference distribution.}$$

**Table EC.1** Number of distributions used in simulation.

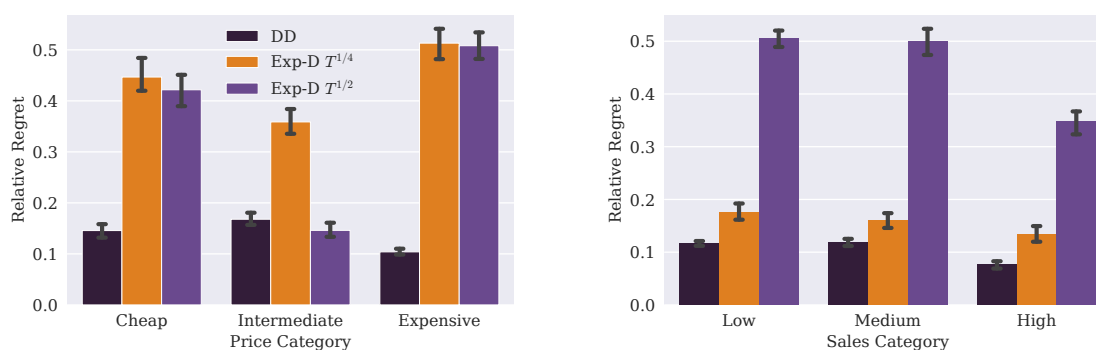
|              | Normal<br>(Nahmias (1994)) | Truncated<br>Normal | Normal | Gamma | Weibull | Chi | Exponential |
|--------------|----------------------------|---------------------|--------|-------|---------|-----|-------------|
| Price Groups | 105                        | 75                  | 20     | 91    | 0       | 7   | 2           |
| Sales Groups | 114                        | 20                  | 62     | 58    | 8       | 21  | 17          |

When multiple distributions are not rejected, we opted for the one with the higher p-value. Moreover, we used Poisson distribution in Section EC.7.4 for testing the Discrete Demand  $\alpha$ -Search algorithm. The selected distributions for the experiment presented in Section 6.2 are given in Table EC.1.

Using the estimated distribution, we simulated demand for 100 periods. This set of experiments runs in 100 random seeds. We used the average price as a base for our parameters. The holding cost is calculated as a percentage of the purchase cost. For the lost sales cost, we used the difference between price and purchase cost. The transportation cost is computed based on the distance between the distribution center and stores. We assumed the salvage cost as zero.

#### EC.7.4. Real-data Experiment with Discrete Demand Distribution

In this section, we repeat the real-data experiment presented in Section 6.2 under discrete demand distribution. We use the same cost structure and estimate Poisson demand. Instead of the DBS algorithm, we used its discrete demand distribution version called DD algorithm 4 with  $\alpha$  search. Also, we modified the Exp algorithm for the discrete demand and obtained the Exp-D algorithm.



**Figure EC.5** Real data experiment with discrete demand. On the left, products are grouped by their prices. On the right, products are grouped by their sales volume.

The results from the experiments are given in Figure EC.5. These results are similar to the ones presented in Section 6.2 in the sense that the DD algorithm performs better than Exp-D algorithms. This suggests that similar to the case under continuous demand distribution, and

the DD algorithm is able to adjust the learning period optimally based on the cost and demand characteristics. As such, the DD algorithm yields more robust performance with respect to changes in business parameters.