Online Appendix to "Technical Note—A Simple Heuristic Policy for Stochastic Distribution Inventory Systems with Fixed Costs"

A. Technical Proofs

In this section, we present technical proofs of some important theoretical results. The proof of other results can be found in Zhu et al. (2020).

Proof of Lemma 3. Let $\hat{\Gamma}_i(IP_i(t))$ be the expected cost rate at Retailer i, with the fixed costs of Type II irregular shipments excluded, when the inventory position at Retailer i is $IP_i(t)$. We first construct an upper bound for $\hat{\Gamma}_i(IP_i(t))$ as follows.

- If $IP_i(t) > r_i$, then Retailer i is in either a regular or irregular shipment interval. To obtain an upper bound on $\hat{\Gamma}_i(IP_i(t))$, we charge the larger one between expected cost rates of the regular and irregular shipment intervals, i.e., $\hat{\Gamma}_i(IP_i(t)) \leq \max\{G_i(IP_i(t)), C_i(r_i, Q_i)\}$. Moreover, by the definition of w_i, it follows that $\max\{G_i(IP_i(t)), C_i(r_i, Q_i)\} \leq \max\{G_i(w_i), C_i(r_i, Q_i)\}$. Therefore, if $IP_i(t) > r_i$, $\hat{\Gamma}_i(IP_i(t)) \leq \max\{G_i(w_i), C_i(r_i, Q_i)\}$.
- If $IP_i(t) \leq r_i$, which implies Retailer i must be in an irregular shipment interval, then we have $\hat{\Gamma}_i(IP_i(t)) = G_i(IP_i(t))$.

Next, we link the upper bound with the echelon inventory level of the warehouse. Obviously, $\hat{\Gamma}(IL_0(t)) = \sum_{i=1}^N \hat{\Gamma}_i(IP_i(t))$. Let $OI_0^+(t)$ denote the on-hand inventory at the warehouse at time t. It follows that $IL_0(t) = OI_0^+(t) + \sum_{j=1}^N IP_j(t)$. Under the modified echelon (r, Q) policy, we have $IP_j(t) \leq r_j + Q_j$ for any $j \in [N]^+$. It follows that

$$OI_0^+(t) + IP_i(t) = IL_0(t) - \sum_{j \neq i} IP_j(t) \geq IL_0(t) - \sum_{j \neq i} (r_j + Q_j). \tag{1}$$

If $OI_0^+(t) > 0$, by the definition the modified echelon (r, Q) policy, we have $r_i < IP_i(t) \leq r_i + Q_i$ for any $i \in [N]$. Then as mentioned above, $\hat{\Gamma}_i(IP_i(t)) \leq \max\{G_i(w_i), C_i(r_i, Q_i)\}$, and thus $\hat{\Gamma}(IL_0(t)) = \sum_{i=1}^N \hat{\Gamma}_i(IP_i(t)) \leq \sum_{i=1}^N \max\{G_i(w_i), C_i(r_i, Q_i)\}$. It is easy to see that (4) always holds.

We next focus on the case $OI_0^+(t) = 0$. Then (1) can be rewritten as

$$IP_i(t) = IL_0(t) - \sum_{j \neq i} IP_j(t) \geq IL_0(t) - \sum_{j \neq i} (r_j + Q_j). \tag{2}$$

Note that if $IP_i(t) \leq r_i$, it follows that $IL_0(t) - \sum_{j \neq i}(r_j + Q_j) \leq IP_i(t) \leq r_i$. Consequently, we have $G_i(IP_i(t)) \leq \max\{G_i(IL_0(t) - \sum_{j \neq i}(r_j + Q_j)), G_i(r_i)\} \leq \max\{G_i(IL_0(t) - \sum_{j \neq i}(r_j + Q_j)), G_i(w_i)\}$ due to the convexity of $G_i(\cdot)$. Therefore, the upper bound of $\hat{\Gamma}_i(IP_i(t))$ is given as follows.

$$\hat{\Gamma}_i(IP_i(t)) \leq \begin{cases} \max\{G_i(w_i), C_i(r_i,Q_i)\} & \text{if } IL_0(t) - \sum_{j\neq i}(r_j+Q_j) > r_i, \\ \max\{G_i(IL_0(t) - \sum_{j\neq i}(r_j+Q_j)), G_i(w_i), C_i(r_i,Q_i)\} & \text{otherwise.} \end{cases} \tag{3}$$

We prove the desired result (4) by induction. We shall start from the case $N=2$.

- If both retailers' inventory position is larger than r_i, i.e., $IP_i(t) > r_i$ for $i=1,2$, we charge $\max\{G_i(w_i), C_i(r_i,Q_i)\}$ for both retailers. Then, the result (4) clearly holds.

- Suppose only one retailer's inventory position is larger than r_i. If $IP_1(t) > r_1$ and $IP_2(t) \leq r_2$, by (3), we charge $\max\{G_1(w_1), C_1(r_1,Q_1)\}$ for Retailer 1, and $\max\{G_2(IL_0(t) - r_1 - Q_1), G_2(w_2), C_2(r_2,Q_2)\}$ for Retailer 2. That is, $\hat{\Gamma}(IL_0(t)) \leq \max\{G_1(w_1), C_1(r_1,Q_1)\} + \max\{G_2(IL_0(t) - r_1 - Q_1), G_2(w_2), C_2(r_2,Q_2)\} = \bar{\Gamma}_2(IL_0(t))$. Similarly, if $IP_1(t) \leq r_1$ and $IP_2(t) > r_2$, one can show that $\hat{\Gamma}(IL_0(t)) \leq \bar{\Gamma}_1(IL_0(t))$. Therefore, $\hat{\Gamma}(IL_0(t)) \leq \max_{i=1,2} \bar{\Gamma}_i(IL_0(t))$.

- Third, suppose both retailers' inventory position is no larger than r_i, i.e., $IP_i(t) \leq r_i$ for $i=1,2$. Then we charge $G_i(IP_1(t))$ for both retailers, and thus we have

$$\begin{aligned} &\hat{\Gamma}(IL_0(t)) = \sum_{i=1}^{2} \hat{\Gamma}_i(IP_i(t)) = G_1(IP_1(t)) + G_2(IL_0(t) - IP_1(t)) \equiv f_2(IP_1(t)) \\ &\leq \max\{G_1(r_1+Q_1) + G_2(IL_0(t) - r_1 - Q_1), G_1(IL_0(t) - r_2 - Q_2) + G_2(r_2+Q_2)\} = \max_{i=1,2} \bar{\Gamma}_i(IL_0(t)), \end{aligned}$$

where the inequality is due to the convexity of $f_2(\cdot)$ and $IL_0(t) - r_2 - Q_2 \leq IP_1(t) \leq r_1 + Q_1$. Therefore, the result (4) holds for $N=2$.

We assume (4) holds for $N=k$. That is,

$$\begin{aligned} &\hat{\Gamma}(\sum_{i=1}^{k} IP_i(t)) = \sum_{i=1}^{k} \hat{\Gamma}_i(IP_i(t)) = \hat{\Gamma}_1(IP_i(t)) + \hat{\Gamma}_2(IP_i(t)) + \cdots + \hat{\Gamma}_{k-1}(IP_{k-1}(t)) + \hat{\Gamma}_k(\sum_{i=1}^{k} IP_i(t) - \sum_{i=1}^{k-1} IP_i(t)) \\ &\leq \max_{i=1,2,\ldots,k} \bar{\Gamma}_i^k(\sum_{i=1}^{k} IP_i(t)), \end{aligned} \tag{4}$$

where $\bar{\Gamma}_i^k(\sum_{i=1}^{k} IP_i(t))$ denotes $\bar{\Gamma}_i(IL_0(t))$, as defined in Lemma 3, in case of $N=k$, i.e.,

$$\begin{aligned} \bar{\Gamma}_i^k(\sum_{i=1}^{k} IP_i(t)) = &\sum_{j\neq i, j\leq k} \max\{G_j(w_j), C_j(r_j,Q_j)\} + \\ &\begin{cases} \max\{G_i(w_i), C_i(r_i,Q_i)\} & \text{if } \sum_{i=1}^{k} IP_i(t) - \sum_{j\neq i, j\leq k}(r_j+Q_j) > r_i, \\ \max\{G_i(\sum_{i=1}^{k} IP_i(t) - \sum_{j\neq i, j\leq k}(r_j+Q_j)), G_i(w_i), C_i(r_i,Q_i)\} & \text{otherwise.} \end{cases} \end{aligned} \tag{5}$$

We next show (4) also holds for $N=k+1$. For ease of statement, we first introduce the expression $\bar{\Gamma}_i(IL_0(t))$ in case of $N=k+1$, denoted as $\bar{\Gamma}_i^{k+1}(IL_0(t))$, as follows.

$$\bar{\Gamma}_i^{k+1}(IL_0(t)) = \sum_{j\neq i, j\leq k+1} \max\{G_j(w_j), C_j(r_j,Q_j)\} + \begin{cases} \max\{G_i(w_i), C_i(r_i,Q_i)\} & \text{if } IL_0(t) - \sum_{j\neq i, j\leq k+1}(r_j+Q_j) > r_i, \\ \max\{G_i(IL_0(t) - \sum_{j\neq i, j\leq k+1}(r_j+Q_j)), G_i(w_i), C_i(r_i,Q_i)\} & \text{otherwise.} \end{cases} \tag{6}$$

It follows that

$$\bar{\Gamma}_i^{k+1}(IL_0(t)) = \bar{\Gamma}_i^{k}(IL_0(t) - r_{k+1} - Q_{k+1}) + \max\{G_{k+1}(w_{k+1}), C_{k+1}(r_{k+1},Q_{k+1})\} \tag{7}$$

Now we are ready to show the desired results.

$$\begin{aligned}
\hat{\Gamma}(IL_0(t)) &= \sum_{i=1}^{k} \hat{\Gamma}_i(IP_i(t)) + \hat{\Gamma}_{k+1}(IL_0(t) - \sum_{i=1}^{k} IP_i(t)) \\
&\leq \max_{i=1,2,\ldots,k} \bar{\Gamma}_i^{k}(\sum_{i=1}^{k} IP_i(t)) + \hat{\Gamma}_{k+1}(IL_0(t) - \sum_{i=1}^{k} IP_i(t)) \equiv f_{k+1}(\sum_{i=1}^{k} IP_i(t)), \\
&\leq \max\{\max_{i=1,2,\ldots,k} \bar{\Gamma}_i^{k}(IL_0(t) - r_{k+1} - Q_{k+1}) + \hat{\Gamma}_{k+1}(r_{k+1}+Q_{k+1}), \\
&\qquad \max_{i=1,2,\ldots,k} \bar{\Gamma}_i^{k}(\sum_{i=1}^{k} r_i + Q_i) + \hat{\Gamma}_{k+1}(IL_0(t) - \sum_{i=1}^{k}(r_i+Q_i))\} \\
&\leq \max\{\max_{i=1,2,\ldots,k} \bar{\Gamma}_i^{k}(IL_0(t) - r_{k+1} - Q_{k+1}) + \max\{G_{k+1}(w_{k+1}), C_{k+1}(r_{k+1},Q_{k+1})\}, \\
&\qquad \max_{i=1,2,\ldots,k} \bar{\Gamma}_i^{k}(\sum_{i=1}^{k} r_i + Q_i) + \hat{\Gamma}_{k+1}(IL_0(t) - \sum_{i=1}^{k}(r_i+Q_i))\} \\
&= \max\{\max_{i=1,2,\ldots,k} \bar{\Gamma}_i^{k+1}(IL_0(t)), \max_{i=1,2,\ldots,k} \bar{\Gamma}_i^{k}(\sum_{i=1}^{k} r_i + Q_i) + \hat{\Gamma}_{k+1}(IL_0(t) - \sum_{i=1}^{k}(r_i+Q_i))\} \\
&= \max\{\max_{i=1,2,\ldots,k} \bar{\Gamma}_i^{k+1}(IL_0(t)), \sum_{j=1}^{k} \max\{G_j(w_j), C_j(r_j+Q_j)\} + \hat{\Gamma}_{k+1}(IL_0(t) - \sum_{i=1}^{k}(r_i+Q_i))\} \\
&\leq \max\{\max_{i=1,2,\ldots,k} \bar{\Gamma}_i^{k+1}(IL_0(t)), \bar{\Gamma}_{k+1}(IL_0(t))\} = \max_{i=1,2,\ldots,k+1} \bar{\Gamma}_i(IL_0(t))\}.
\end{aligned}$$

The first equality follows from (2). The first inequality follows from (4). The second inequality is due to the convexity of $f_{k+1}(\cdot)$ and $IL_0(t) - r_{k+1} - Q_{k+1} \leq \sum_{i=1}^{k} IP_i(t) \leq \sum_{i=1}^{k}(r_i+Q_i)$. The third inequality is because we charge $\max\{G_{k+1}(w_{k+1}), C_{k+1}(r_{k+1},Q_{k+1})\}$ for Retailer $k+1$ when its inventory position is $r_{k+1}+Q_{k+1}$, i.e., $\hat{\Gamma}_{k+1}(r_{k+1}+Q_{k+1}) \leq \max\{G_{k+1}(w_{k+1}), C_{k+1}(r_{k+1},Q_{k+1})\}$. The second equality follows from (7). The third equality follows from the definition of $\bar{\Gamma}_i^{k}(\cdot)$; see (5). The last inequality holds true due to (3) and $IL_0(t) - \sum_{i=1}^{k}(r_i+Q_i) \leq IP_{k+1}(t) \leq r_{k+1}+Q_{k+1}$. The last equality follows from (6). $\square$

Proof of Theorem 1. By (4) and the definition of $\hat{G}(y)$ in (5), we obtain

$$\hat{\Gamma}(IL_0(t)) \leq \bar{\Gamma}(IL_0(t)) = \hat{G}(IL_0(t)) + \sum_{i=1}^{N} C_i(r_i, Q_i). \tag{8}$$

We denote by $\Gamma_0(IP_0(t))$ the total expected cost rate of all installations at time t when the inventory position of the warehouse is $IP_0(t)$, where we exclude the fixed costs incurred at the warehouse and the fixed costs of Type II irregular shipment intervals incurred at retailers. By such a definition, $\Gamma_0(IP_0(t))$ constitutes two parts: (i) the inventory holding cost at the warehouse, and (ii) the total costs at all retailers excluding the fixed costs in Type II irregular shipment intervals. Then, according to the cost accounting scheme, we have

$$\begin{aligned}\Gamma_0(IP_0(t)) =& \mathbb{E}[h_0(IP_0(t) - D_0)] + \mathbb{E}[\hat{\Gamma}(IP_0(t) - D_0)] \\ \leq& \mathbb{E}[h_0(IP_0(t) - D_0)] + \mathbb{E}[\hat{G}(IP_0(t) - D_0)] + \sum_{i=1}^{N} C_i(r_i, Q_i) = \Lambda_0(IP_0(t)) + \sum_{i=1}^{N} C_i(r_i, Q_i),\end{aligned} \tag{9}$$

where the inequality follows from (8), and the last equality from (6).

Because the warehouse has an unlimited supply form the external supplier, under the modified echelon (r, Q) policy, the inventory position of the warehouse, $IP_0(t)$, is uniformly distributed on $\{r_0 + 1, \ldots, r_0 + Q_0\}$. Therefore, by the definition of $\Gamma_0(IP_0(t))$, the long-run average system-wide cost, with the fixed costs of Type II irregular shipment intervals incurred at retailers being excluded, can be bounded as

$$\begin{aligned}\frac{1}{Q_0}\left[\lambda_0 K_0 + \int_{r_0}^{r_0+Q_0} \Gamma_0(y)dy\right] \leq& \frac{1}{Q_0}\left[\lambda_0 K_0 + \int_{r_0}^{r_0+Q_0} [\Lambda_0(y) + \sum_{i=1}^{N} C_i(r_i, Q_i)]dy\right] \\ =& \hat{C}_0(r_0, Q_0) + \sum_{i=1}^{N} C_i(r_i, Q_i),\end{aligned} \tag{10}$$

where the inequality is due to (9) and the equality holds true due to (7).

Finally, by incorporating the upper bound on the fixed cost of Type II irregular shipment intervals, we can obtain that the long-run average system-wide cost can be bounded as: $C(\mathbf{r}, \mathbf{Q}) \leq \hat{C}_0(r_0, Q_0) + \sum_{i=1}^{N} C_i(r_i, Q_i) + \lambda_0 \overline{K}/Q_0$. $\square$

Proof of Theorem 2. (i) As shown in Lemma 4 in Zhu et al. (2020), the cost lower bound is $\sum_{i=0}^{N} C_i^*$. As shown in (12), the upper bound cost is $\sum_{i=1}^{N} C_i^* + \tilde{C}_0^*$. It follows that the MERQD policy is at least $1+(\tilde{C}_0^* - C_0^*)/(\sum_{i=1}^{N} C_i^* + C_0^*)$-optimal. The last result can be obtained by showing $(\sum_{i=1}^{N} C_i^* + \tilde{C}_0^*)/(\sum_{i=1}^{N} C_i^* + C_0^*) \leq \tilde{C}_0^*/C_0^*$. It suffices to show $(\sum_{i=1}^{N} C_i^* + \tilde{C}_0^*)C_0^* \leq (\sum_{i=1}^{N} C_i^* + C_0^*)\tilde{C}_0^*$. Then, the desired result directly holds because $C_0^* \leq \tilde{C}_0^*$.

(ii) By (8) and the definition of $(\hat{\mathbf{r}}, \hat{\mathbf{Q}})$, we have that for any (r_0, Q_0), $\mathcal{C}^*_{\mathcal{B}} \le C(\hat{\mathbf{r}}, \hat{\mathbf{Q}}) \le \sum_{i=1}^N C_i^* + \hat{C}_0(r_0, Q_0) + \lambda_0 \overline{K}/Q_0$. By Lemmas 1(iii) and 2(ii) in Zhu et al. (2020), we can obtain that for any (r_0, Q_0),

$$\begin{aligned}\mathcal{C}^*_{\mathcal{B}} \le C(\hat{\mathbf{r}}, \hat{\mathbf{Q}}) &\le \sum_{i=1}^N C_i^* + \hat{C}_0(r_0, Q_0) + \frac{\lambda_0 \overline{K}}{Q_0} \\ &\le \sum_{i=1}^N C_i^* + \hat{C}_0(r_0, Q_0) + \frac{\lambda_0 C_m^* Q_m^*}{2\lambda_m Q_0} \\ &\le \sum_{i=1}^N C_i^* + \epsilon(\frac{Q_0}{\hat{Q}_0^*})\hat{C}_0^* + \frac{\lambda_0 C_m^* Q_m^*}{2\lambda_m Q_0}. \end{aligned} \tag{11}$$

Recall that $\hat{C}_0^*$ and $\hat{Q}_0^*$ are the optimal cost and order quantity for $\hat{C}_0(r_0, Q_0)$, respectively. Therefore, the last inequality follows from Lemma 1(iii) in Zhu et al. (2020). We select Q_0 in (11) as

$$\check{Q}_0 \equiv \arg\min_{Q_0} \Big\{ \epsilon(\frac{Q_0}{\hat{Q}_0^*})\hat{C}_0^* + \frac{\lambda_0 C_m^* Q_m^*}{2\lambda_m Q_0} \Big\} = \sqrt{\frac{(\hat{Q}_0^*)^2 \hat{C}_0^* + (\lambda_0 C_m^* Q_m^*)\hat{Q}_0^*/\lambda_m}{\hat{C}_0^*}} = \hat{Q}_0^* \sqrt{1 + \frac{\lambda_0 C_m^* Q_m^*}{\lambda_m \hat{C}_0^* \hat{Q}_0^*}}.$$

By replacing Q_0 in (11) with $\check{Q}_0$, we can obtain a new upper bound: $\mathcal{C}^*_{\mathcal{B}} \le C(\hat{\mathbf{r}}, \hat{\mathbf{Q}}) \le \sum_{i=1}^N C_i^* + \hat{C}_0^* \sqrt{1 + \frac{\lambda_0 C_m^* Q_m^*}{\lambda_m \hat{C}_0^* \hat{Q}_0^*}}$. It follows that the relative gap between $\mathcal{C}^*_{\mathcal{B}}$ and $C(\hat{\mathbf{r}}, \hat{\mathbf{Q}})$ is bounded as

$$\begin{aligned}(C(\hat{\mathbf{r}}, \hat{\mathbf{Q}}) - \mathcal{C}^*_{\mathcal{B}})/\mathcal{C}^*_{\mathcal{B}} \le \hat{C}_0^* \Big(\sqrt{1 + \frac{\lambda_0 C_m^* Q_m^*}{\lambda_m \hat{C}_0^* \hat{Q}_0^*}} - \beta_2 \Big) / \mathcal{C}^*_{\mathcal{B}} &\le \hat{C}_0^* \Big(\sqrt{1 + \frac{\lambda_0 C_m^* Q_m^*}{\lambda_m \hat{C}_0^* \hat{Q}_0^*}} - \beta_2 \Big) / (\beta_2 \hat{C}_0^* + \sum_{i=1}^N C_i^*) \\ &\le \hat{C}_0^* \Big(\sqrt{1 + \frac{\lambda_0 C_m^* Q_m^*}{\lambda_m \hat{C}_0^* \hat{Q}_0^*}} - \beta_2 \Big) / (\beta_2 \hat{C}_0^* + C_m^*).\end{aligned}$$

To obtain the desired result, it is sufficient to show the following stronger statement: for any $x_1, x_2 > 0$, $x_2(\sqrt{1 + \lambda_0 x_1/(\lambda_m x_2 \beta_1)} - \beta_2)/(\beta_2 x_2 + x_1) \le \alpha \equiv \max\{\sqrt{\frac{\lambda_0}{2\beta_1\beta_2\lambda_m} + \frac{1}{4}} - \frac{1}{2}, \frac{1}{\beta_2} - 1\}$, which is equivalent to $\alpha^2 x^2 + [2\beta_2\alpha(1+\alpha) - \lambda_0/(\lambda_m\beta_1)]x + \beta_2^2\alpha(\alpha+2) + \beta_2^2 - 1 \ge 0$ for any $x > 0$. By verifying that (1) $2\beta_2\alpha(1+\alpha) - \lambda_0/(\lambda_m\beta_1) \ge 0$ when $\alpha \ge \sqrt{\frac{\lambda_0}{2\beta_1\beta_2\lambda_m} + \frac{1}{4}} - \frac{1}{2}$; and (2) $\beta_2^2\alpha(\alpha+2) + \beta_2^2 - 1 \ge 0$ when $\alpha \ge \frac{1}{\beta_2} - 1$, the desired result is obtained.

We prove the alternative bound mentioned in Footnote 2. Following part (ii), it suffices to show that the quadratic function $\alpha^2 x^2 + [2\beta_2\alpha(1+\alpha) - \lambda_0/(\lambda_m\beta_1)]x + \beta_2^2\alpha(\alpha+2) + \beta_2^2 - 1 \ge 0$ when $\alpha \ge \lambda_0/(2(\beta_1\beta_2\lambda_m + \sqrt{(\beta_2^2-1)(\beta_1\lambda_m)^2 + \beta_1\beta_2\lambda_m\lambda_0}))$ under the condition $(\beta_2^2-1)(\beta_1\lambda_m) + \beta_2\lambda_0 \ge 0$. The bound is then established by verifying that the quadratic function $f(x) = \alpha^2 x^2 + [2\beta_2\alpha(1+\alpha) - \lambda_0/(\lambda_m\beta_1)]x + \beta_2^2\alpha(\alpha+2) + \beta_2^2 - 1$ has a zero discriminant, i.e., $\Delta = 4\alpha^2(1 - \lambda_0\beta_2/(\lambda_m\beta_1)) - 4\alpha\beta_2\lambda_0/(\lambda_m\beta_1) + (\lambda_0/(\lambda_m\beta_1))^2 = 0$. □

Proof of Theorem 3. To prove asymptotic optimality of the modified echelon (r, Q) policy, it is sufficient to first show the following statements: (i) $\lim_{K_0/K_m\to\infty}\beta_1=\infty$ and $\lim_{K_0/K_m\to\infty}\beta_2=1$; (ii) $\lim_{h_0/h_m\to 0}\beta_1=\infty$ and $\lim_{h_0/h_m\to 0}\beta_2=1$; (iii) $\lim_{h_0/p_m\to 0}\beta_1=\infty$ and $\lim_{h_0/p_m\to 0}\beta_2=1$.

(i) Let $r_i(Q_i)=\arg\min_{r_i} C_i(r_i,Q_i)$ for $i=0,1,2,\dots,N$. Define $A_i(Q_i)\equiv Q_iG_i(r_i(Q_i))-\int_0^{Q_i}G_i(r_i(y))dy$. By Lemma 1 in Zhu et al. (2020), $A_i(Q_i)$ is a continuous and strictly increasing function such that $A_i(Q_i^*)=\lambda_iK_i$ and $A_i(0)=0$. Let $A_i^{-1}(x)$ be the inverse function of $A_i(Q_i)$. Then, $Q_i^*=A_i^{-1}(\lambda_iK_i)$ and $A_i^{-1}(0)=0$. Similarly, let $\hat{r}_0(Q_0)=\arg\min_{r_0}\hat{C}_0(r_0,Q_0)$ and define $\hat{A}_0(Q_0)\equiv Q_0\Lambda_0(\hat{r}_0(Q_0))-\int_0^{Q_0}\Lambda_0(\hat{r}_0(y))dy$. Let $\hat{A}_0^{-1}(x)$ be the inverse function of $\hat{A}_0(Q_0)$. Then, $\hat{Q}_0^*=\hat{A}_0^{-1}(\lambda_0K_0)$ and $\hat{A}_0^{-1}(0)=0$. Let $\gamma=K_m/K_0$. We have $\lim_{K_0/K_m\to\infty}\beta_1=\lim_{K_0/K_m\to\infty}\frac{\hat{A}_0^{-1}(\lambda_0K_0)}{A_m^{-1}(\lambda_mK_m)}=\lim_{\gamma\to 0}\frac{\hat{A}_0^{-1}(\lambda_0K_0)}{A_m^{-1}(\lambda_m\gamma K_0)}=\frac{\hat{A}_0^{-1}(\lambda_0K_0)}{A_m^{-1}(\lambda_mK_0)}\lim_{\gamma\to 0}\frac{A_m^{-1}(\lambda_mK_0)}{A_m^{-1}(\lambda_m\gamma K_0)}=\infty$, where the last equality follows from the fact that $A_m^{-1}(x)$ is continuous and $A_m^{-1}(0)=0$.

It remains to show $\lim_{K_0/K_m\to\infty}\beta_2=1$. Recall that $A_0(Q)$ and $\hat{A}_0(Q)$ are both increasing convex functions. Let $r(Q_0)=\arg\min_{r_0}C_0(r_0,Q_0)$. Define $H_0(Q)\equiv G_0(r(Q))$ and $\hat{H}_0(Q)\equiv\Lambda_0(\hat{r}_0(Q))$. It follows that $A_0'(Q)=QH_0(Q)$ and $\hat{A}_0'=Q\hat{H}_0(Q)$. By Lemma 3(i) in Zhu et al. (2020), we have

$$\lim_{Q\to\infty}H_0'(Q)=\frac{h_0\underline{p}}{h_0+\underline{p}},\qquad \lim_{Q\to\infty}\hat{H}_0'(Q)=\frac{h_0\overline{p}}{h_0+\overline{p}}, \tag{12}$$

where $\underline{p}=\min_{i\in[N]^+}\{p_i\}$. With the above analysis, under the condition $K_m>0$, we can obtain

$$\begin{aligned}
\lim_{K_0/K_m\to\infty}\beta_2&=\lim_{\gamma\to 0}\frac{H_0[A_0^{-1}(\lambda_0K_m/\gamma)]}{\hat{H}_0[\hat{A}_0^{-1}(\lambda_0K_m/\gamma)]}\\
&=\lim_{\gamma\to 0}\frac{H_0'[A_0^{-1}(\lambda_0K_m/\gamma)]\frac{dA_0^{-1}(\lambda_0K_m/\gamma)}{d\gamma}}{\hat{H}_0'[\hat{A}_0^{-1}(\lambda_0K_m/\gamma)]\frac{d\hat{A}_0^{-1}(\lambda_0K_m/\gamma)}{d\gamma}}\\
&=\lim_{\gamma\to 0}\frac{H_0'[A_0^{-1}(\lambda_0K_m/\gamma)]}{\hat{H}_0'[\hat{A}_0^{-1}(\lambda_0K_m/\gamma)]}\lim_{Q\to\infty}\frac{\hat{A}_0'(Q)}{A_0'(Q)}\\
&=\lim_{\gamma\to 0}\frac{H_0'[A_0^{-1}(\lambda_0K_m/\gamma)]}{\hat{H}_0'[\hat{A}_0^{-1}(\lambda_0K_m/\gamma)]}\lim_{Q\to\infty}\frac{Q\hat{H}_0'(Q)}{QH_0'(Q)}=1,
\end{aligned}$$

where the second equality is due to the L'Hospital's Rule, the third from the rules for taking derivative of inverse functions, and the last from (12).

(ii) Let $\xi\equiv h_0/h_m$. Then $G_m(y)$ can be written as $G_m(y)=\mathbb{E}[h_m(y-D_m(t,t+L_m])+(h_m+\xi h_m+p_m)(y-D_m(t,t+L_m])^-]$. It follows that for any $p_m>0$ and $h_m>0$, Q_m^* converges to a constant when $\xi\to 0$. Then we proceed to show that $\lim_{\xi\to 0}\hat{Q}_0^*=\infty$. Let $\bar{H}_0(Q)\equiv\frac{h_0\overline{p}}{h_0+\overline{p}}Q$, $\bar{A}_0(Q)\equiv\frac{h_0\overline{p}}{2(h_0+\overline{p})}Q^2$, and $\bar{Q}_0\equiv\sqrt{\frac{2\lambda_0K_0(h_0+\overline{p})}{h_0\overline{p}}}$. Then, it is easy to check that $\bar{A}_0'=Q\bar{H}_0'(Q)$ and $\bar{A}_0(\bar{Q}_0)=\lambda_0K_0$. Because $\hat{H}_0(Q)$ is an increasing convex functions, by (12), we have $\hat{H}_0'(Q)\le\frac{h_0\overline{p}}{h_0+\overline{p}}=\bar{H}_0'(Q)$, which implies that $\hat{A}_0'(Q)\le\bar{A}_0'(Q)$. Note that $\hat{A}_0(0)=\bar{A}_0(0)=0$. Therefore, we must have $\hat{A}_0(Q)\le\bar{A}_0(Q)$. We then have $\hat{A}_0(\bar{Q}_0) \le \bar{A}_0(\bar{Q}_0) = \lambda_0 K_0$. Because $\hat{A}_0(\hat{Q}_0^*) = \lambda_0 K_0$ and $\hat{A}_0(Q)$ is an increasing function, we have $\hat{Q}_0^* \ge \bar{Q}_0$. On the other hand, because $K_0 > 0$ and $\lim_{\xi\to 0}\bar{Q}_0 = \sqrt{\frac{2\lambda_0 K_0(\xi h_m + \bar{p})}{\xi h_m \bar{p}}} = \infty$, we have $\lim_{\xi\to 0}\hat{Q}_0^* = \infty$. Therefore, we have $\lim_{\xi\to 0}\beta_1 = \infty$.

It remains to show $\lim_{\xi\to 0}\beta_2 = 1$. Recall that we always have $C_0^* \le \hat{C}_0^* \le \hat{C}_0(r_0^*, Q_0^*)$, where the second equality follows from the definition of $\hat{C}_0^*$. Therefore, it suffices to show $\lim_{\xi\to 0}\hat{C}_0(r_0^*, Q_0^*) = \lim_{\xi\to 0} C_0^*$. Following the same logic of showing $\lim_{\xi\to 0}\hat{Q}_0^* = \infty$, one can easily prove $\lim_{\xi\to 0} Q_0^* = \infty$. Then we have $\lim_{\xi\to 0}\frac{C_0^*}{\hat{C}_0(r_0^*,Q_0^*)} = \lim_{\xi\to 0}\frac{H_0(Q_0^*)}{\hat{H}_0(Q_0^*)} = \lim_{\xi\to 0}\frac{H_0'(Q_0^*)}{\hat{H}_0'(Q_0^*)} = \lim_{\xi\to 0}\frac{\frac{\xi h_m p}{\xi h_m + p}}{\frac{\xi h_m[\sum_{i=1}^N (p_i+\xi h_m) - \xi h_m]}{\sum_{i=1}^N (p_i + \xi h_m)}} = 1$, where the second equality holds by the L'Hospital's Rule, and the third holds due to (12).

(iii) The proof emulates that of (ii). Let $\mu \equiv h_0/p_m$. Then $G_m(y)$ can be written as $G_m(y) = \mathbb{E}[h_m(y - D_m(t, t+L_m]) + (h_m + \mu p_m + p_m)(y - D_m(t, t+L_m])^-]$. It follows that for any $p_m > 0$ and $h_m > 0$, Q_m^* converges to a constant when $\mu \to 0$. In addition, in this case we have $\lim_{\mu\to 0}\bar{Q}_0 = \sqrt{\frac{2\lambda_0 K_0(\mu p_m + \bar{p})}{\mu p_m \bar{p}}} = \infty$. Therefore, we have $\lim_{\mu\to 0}\hat{Q}_0^* = \infty$ and $\lim_{\mu\to 0}\beta_1 = \infty$. To see $\lim_{\mu\to 0}\beta_2 = 1$, note that in this case we have $\lim_{\mu\to 0}\frac{C_0^*}{\hat{C}_0(r_0^*,Q_0^*)} = \lim_{\mu\to 0}\frac{\frac{\mu p_m p}{\mu p_m + p}}{\frac{\mu p_m \bar{p}}{\mu p_m + \bar{p}}} = 1$. □

B. Numerical Experiments

B.1. Overall Performance

Although we can efficiently compute the cost upper bound of our heuristic (see Theorem 1), the exact computation of the real cost of the heuristic is technically challenging. To evaluate the exact performance, we use the Monte Carlo simulation method to compute the long-run average cost of the inventory system under this heuristic policy. We denote by $\tilde{C}$ the real cost obtained by the Monte Carlo simulation, which is a sample mean with a sample error less than 0.05 and a 95% confidence interval. Unless otherwise mentioned, $\tilde{C}$ is computed under the inventory position priority rule in which retailers are replenished based on the reverse order of their inventory position. We define the following percentage:

$$\delta_1 \equiv \frac{\tilde{C} - \underline{C}^*}{\underline{C}^*} \times 100\%,$$

which is an upper bound on the performance gap of our heuristic because we benchmark it with the cost lower bound. The complete test set of primitive values is given by $L_0 \in \{0,1,2\}$, $L_1 \in \{0,1,2\}$, $K_0 \in \{100, 200, 600\}$, $K_1 \in \{10, 20, 40\}$, $h_0 \in \{0.05, 0.1, 0.2\}$, $h_1 \in \{0.3, 0.5, 1\}$, $p_1 \in \{3, 5, 10\}$, and $\lambda_1 \in \{3, 5, 7\}$, with the other primitives fixed as $N = 2$, $L_2 = 1$, $K_2 = 20$, $h_2 = 0.5$, $p_2 = 5$, and $\lambda_2 = 5$. All combinations of these primitives provide $3^8 = 6561$ test instances.

The numerical results are summarized in Table 1. The average gap δ_1 between the cost lower

	Average (%)	Standard deviation (%)	Minimum (%)	Maximum (%)
δ_1	5.76	3.07	1.45	20.30
δ_1	$\leq 5\%$	$\leq 10\%$	$\leq 20\%$	$\leq 20.3\%$
Number of instances	3263	5833	6558	6561
Cumulative percent (%)	49.73	88.90	99.95	100

Table 1 Overview of the performance of the modified echelon $(\hat{r}, \hat{Q})$ policy.

bound and the cost of the MERQD policy is around 5.76%, with a minimum of about 1.45%. Table 1 shows that the maximum δ_1 in our test is no more than 21%, i.e., the MERQD policy can guarantee at least 1.21-optimality for our test instances. Moreover, the gap δ_1 is less than 10% in about 90% instances, and less than 20% in 99.95% instances. Because the optimal policy is unknown and we benchmark the performance of our heuristic against the induced penalty lower bound, our heuristic may perform even better than the results reported in Table 1. For discussions on the effectiveness of the lower bound, see Axsäter et al. (2002), Gallego et al. (2007), and Doğru et al. (2009).

We relegate the detailed numerical experiments to an online supplement Zhu et al. (2020). Specifically, through extensive numerical studies, we investigate the impacts of system primitives on the performance of our heuristic, compare the performance of the MERQD policy with that of the echelon-stock (r, Q) policy used in Chen and Zheng (1997), and test the robustness of the MERQD policy with respect to the allocation rule at the warehouse.

References

Axsäter S, Marklund J, Silver EA (2002) Heuristic methods for centralized control of one-warehouse, n-retailer inventory systems. *Manufacturing & Service Oper. Management* 4(1):75–97.

Chen F, Zheng YS (1997) One-warehouse multiretailer systems with centralized stock information. *Oper. Res.* 45(2):275–287.

Doğru MK, De Kok A, Van Houtum G (2009) A numerical study on the effect of the balance assumption in one-warehouse multi-retailer inventory systems. *Flex. Serv. Manuf. J.* 21(3-4):114–147.

Gallego G, Özer Ö, Zipkin P (2007) Bounds, heuristics, and approximations for distribution systems. *Oper. Res.* 55(3):503–517.

Zheng YS (1992) On properties of stochastic inventory systems. *Management Sci.* 38(1):87–103.

Zhu H, Chen F, Hu M, Yang Y (2020) Supplement to "a simple heuristic policy for stochastic distribution inventory systems with fixed shipment costs". `ssrn.com/abstract=3606084`, online supplement.