E-Companion for "Online Learning for Dual Index Policies in Dual Sourcing Systems"

Appendix A: Summary of Major Notation

Table 2: Summary of Major Notation for Model Formulation

T	the total number of periods
D^t	the demand in period t, random variable
D	time generic demand variable
d^t	the realized demand in period t
D_k^t	the cumulative demand from period t to $t+k$, $D_k^t = \sum_{i=0}^{k} D^{t+i}$, $\forall t \in [T], k \in \mathbb{Z}$
d_k^t	the realization of D_k^t
c_e	the unit ordering cost of inventory from expedited channel
c_r	the unit ordering cost of inventory from regular channel
l_e	the lead time of inventory from expedited channel
l_r	the lead time of inventory from regular channel
l	$l_r - l_e \geq 1$, the difference between two lead times
h	the unit holding cost for excess inventory
b	the unit penalty cost for unmet demand
IP_e^t	the expedited inventory position in period t
IP_r^t	the regular inventory position in period t
z_e	the order-up-to level for expedited inventory position
z_r	the order-up-to level for regular inventory position
Δ	$z_r - z_e$, the difference between the order-up-to levels
q_e^t	the order in period t from expedited channel
q_r^t	the order in period t from regular channel
I^t	the on-hand inventory in period t
O^t	the overshoot in period t, random variable
o^t	the realized overshoot in period t

Table 3: Summary of Major Notation for Objective and Regret

$W^t(z_e, z_r)$	$(q_r^{t-1}, \ldots, q_r^{t-l+1}, IP_e^t + q_r^{t-l}) \in \mathbb{R}_+^{l-1} \times \mathbb{R}$ state variable under dual-index policy z_e, z_r
$C^t(z_e, z_r)$	the cost in period t under dual-index policy (z_e, z_r)
$\bar{W}^t(z_e, z_r)$	$(q_r^{t-1}, \ldots, q_r^{t-l+1}, 0, \ldots, 0, \max(z_e, IP_e^t + q_r^{t-l})) \in \mathbb{R}_+^l$
$O^\infty(\Delta)$	the random variable following the steady state distribution of the overshoot $O^t = (IP_e^t + q_r^{t-l} - z_e)^+$
$W^\infty(z_e, z_r)$	the steady state of process $W^t(z_e, z_r)$ under dual-index policy z_e, z_r
$C^\infty(z_e, z_r)$	the cost per period in the steady state $W^\infty(z_e, z_r)$
D_{l_e}	the sum of $l_e + 1$ random variable D
ALG	an algorithm for the inventory replenishment in the dual-sourcing system
$C_{\mathbf{ALG}}^t$	the revenue in period t of an algorithm **ALG**
$\mathcal{R}_T^{\mathbf{ALG}}$	the regret of an algorithm **ALG** in T periods
z_e^*	the expedited order-up-to level in the clairvoyant dual-index policy
z_r^*	the regular order-up-to level in the clairvoyant dual-index policy
Δ^*	the gap between the regular and the expedited order-up-to levels in the clairvoyant dual-index policy
$\gamma(z_e, z_r)$	$\mathbb{P}\left(D \leq \dfrac{z_r - z_e}{l_r + 1}\right)$
$\underline{\Delta}$	the lower limit of the gap between the regular and expedited order-up-to level Δ
λ	a known upper bound of $\mathbb{P}\left(D \leq \frac{\bar{z}}{l_r+1}\right)$ with $\lambda < 1$

Table 4: Summary of Major Notation for Online Learning Algorithm

$\bar{D}$	the upper bound of the demand variable
$\bar{Z}$	the upper limit of the regular order-up-to level z_r
μ	the expectation of the demand D
$\underline{\mu}$	the lower limit of the demand mean μ
B^n	the length of each epoch n
L^n	$L^n = \sum_{i=1}^n B^i$, $\forall n \in [N-1]$ with $L^0 = 0, L^N = T$
N	the total number of epochs, $N = \min\left\{n : \sum_{i=1}^n \lceil \frac{2^i}{\log T} \rceil \geq T\right\}$
J	the number of discretized values for Δ in the algorithm, $J = \lfloor \sqrt{T} \rfloor$
F_Δ	the CDF of the distribution of variable $D_{l_e} - O^\infty(\Delta)$
$z_e^*(\Delta)$	the optimal expedited order-up-to level given Δ
$\mathcal{A}^n$	the active set of choices for Δ in epoch n
j^n	the index in $\mathcal{A}^n$ of the Δ selected in epoch n
$\mathcal{D}^n$	the demand data set in epoch n
z_{ej}^n	the estimated optimal expedited order-up-to level given Δ_j in epoch n
$\hat{W}_j^t$	the simulated process for using Δ_j and z_{ej}^n where $t \in [L^n]$
$\hat{G}_j^n$	the estimated average period cost for the dual-index policy with Δ_j and z_{ej}^n in epoch n
$\mathcal{X}_j^n$	the data sample set of $X_j^t = D_{-l_e}^t - O^{t-l_e}(\Delta_j)$ for estimating the empirical quantile in epoch n
$\hat{F}_{\Delta_j}^n(\cdot)$	the empirical CDF of variable $D_{l_e} - O^\infty(\Delta_j)$ using $\mathcal{X}_j^n$
ε^n	the error bound to prune the active set for Δ
T_0	a constant defined in (9)

Table 5: Summary of Major Notation for Regret Analysis

α^n, β^n	parameters in the algorithm
$\bar{C}$	an upper limit of the cost per period $(c_e + c_r + h)\bar{Z} + b((l_e+1)\bar{D})$
K_0	constant, which is $e^{\frac{4(\bar{w}_j^1 \cdot \mathbf{1}^{l-z_r})}{\bar{D}} + \frac{2\mu^2 l_r}{\bar{D}^2}}$
π	our proposed learning algorithm (Δ, z_e)
C_π^t	the cost in period t by running our algorithm $\pi = (\Delta, z_e)$
$\mathcal{R}_T^\pi$	the regret of our learning algorithm π in T periods
$C^t(\Delta, z_e)$	the cost in period t under dual-index policy $(z_e, z_e + \Delta)$
$C^\infty(\Delta, z_e)$	the steady-state per-period cost under dual-index policy $(z_e, z_e + \Delta)$
S	the event that the inventory position drops down below z_r after $\lceil \frac{5\bar{D}^2}{4\mu^2} \log T \rceil$ periods
τ	constant defined as $\tau = \lceil \frac{5\bar{D}^2}{4\mu^2} \log T \rceil + 2l_r \lceil 5 (\log T)^2 \rceil$
U	the event that the demand pattern (4) occurs during periods $\lceil \frac{5\bar{D}^2}{4\mu^2} \log T \rceil + 1$ to period $\lceil \frac{5\bar{D}^2}{4\mu^2} \log T \rceil + 2l_r \lceil 5 (\log T)^2 \rceil$
V^n	the event that two processes $W^t(z_e, z_r) = \tilde{W}^t(z_e, z_r)$ couple after τ periods in epoch n
M_j^n	the event that the estimated cost for arm j in epoch n is accurate enough
N_0	constant defined as $N_0 = \log_2 \log T + \log_2 \left(10 l_r (\log T)^2 + \frac{5\bar{D}^2}{4\mu^2} \log T + 2l_r + 1\right)$
A_j^n	the event that $A_j^n = \left\{ \left\lvert F_{\Delta_j}\left(z_{e_j}^n\right) - \frac{b}{b+h} \right\rvert \leq \alpha^n \right\}$

Appendix B: Proof of Theorem 1

We prove the ergodicity in two disjoint cases depending on the initial regular inventory position.

B.1. Case 1: initial regular inventory position is at most z_r

LEMMA 14. *If* $\gamma(z_e, z_r) = \mathbb{P}\left(D \leq \dfrac{z_r - z_e}{l_r + 1}\right) > 0$, *then the Markov chain* $\{W^t(z_e, z_r) : t \geq 1\}$ *is ergodic with a steady state random vector* $W^\infty(z_e, z_r)$. *Moreover, for any* $t \geq 2l_r + 1$, *any initial vector* $w^1 \in \mathbb{R}_+^{l-1} \times \mathbb{R}$

satisfying $\bar{w}^1 \cdot \mathbf{1}^l \le z_r$,

$$\delta^{t+1}\left(z_e, z_r, w^1\right) \le \left(1-\gamma(z_e,z_r)^{2l_r}\right)^{t/2l_r},$$

where we define $\bar{W}^t(z_e,z_r) = (q_r^{t-1},\dots,q_r^{t-l+1},0,\dots,0,\max(z_e, IP_e^t + q_r^{t-l})) \in \mathbb{R}_+^{l-1}\times\mathbb{R}$.

Proof of Lemma 14. We say a measurable set $\bar{U} \subseteq \mathbb{R}_+^{l-1}\times\mathbb{R}$ is a small set with respect to a nontrivial measure ν provided that there exists $t^*>0$ such that for any $w^1 \in \bar{U}$ and any measurable set $\Omega \subseteq \mathbb{R}_+^{l-1}\times\mathbb{R}$,

$$\mathbb{P}\left(W^{t^*}(z_e,z_r) \in \Omega \,|\, W^1(z_e,z_r) = w^1\right) \ge \nu(\Omega).$$

The following result appears in Theorem 16.0.2 in Meyn and Tweedie (1993). If $\bar{U}$ is a small set with respect to ν, then there exists stationary random variable $W^\infty(z_e,z_r)$ such that for any $w^1 \in \bar{U}$ and $t \ge t^*$,

$$\delta^{t+1}\left(z_e, z_r, w^1\right) \le \left(1-\nu(\mathbb{R}_+^{l-1}\times\mathbb{R})\right)^{t/(t^*-1)}.$$

Recall that $\bar{W}^t(z_e,z_r) = (q_r^{t-1},\dots,q_r^{t-l+1},0,\dots,0,\max(z_e, IP_e^t + q_r^{t-l})) \in \mathbb{R}_+^{l-1}\times\mathbb{R}$. Now we let $\bar{U} = \left\{w^1 \in \mathbb{R}_+^{l-1}\times\mathbb{R} | \bar{w}^1\cdot\mathbf{1} \le z_r\right\}$.

For any $0 \le k \le l_r - 1$, let $\Omega_k \subseteq \mathbb{R}_+$ be any measurable set and let

$$\Omega = \left\{ (q_r^{-1}, q_r^{-2},\dots,q_r^{-l+1}, IP_e + q_r^{-l}) \in \mathbb{R}_+^{l-1}\times\mathbb{R} \middle| q_r^{-k} \in \Omega_k,\ \forall 1 \le k \le l-1, z_r - IP_e - q_r^{-l} - \sum_{k=1}^{l-1} q_r^{-k} \in \Omega_0 \right\}.$$

Define measure

$$\nu(\Omega) = \gamma(z_e,z_r)^{l_r+l_e} \cdot \prod_{k=0}^{l-1} \mathbb{P}\left(D \in \Omega_k \cap \left[0, \frac{z_r - z_e}{l_r+1}\right]\right).$$

So we have $\nu(\mathbb{R}_+^{l-1}\times\mathbb{R}) = \gamma(z_e,z_r)^{2l_r} > 0$. So ν is a non-trivial measure.

To show $\bar{U}$ is a small set with respect to ν and $t^* = 2l_r+1$, we define $\hat{\Omega}_k = \Omega_k \cap \left[0, \frac{z_r-z_e}{l_r+1}\right]$, $\forall 1 \le k \le l-1$. So we have

$$\mathbb{P}\left(W^{2l_r+1}(z_e,z_r) \in \Omega | W^1(z_e,z_r) = w^1\right) \ge \mathbb{P}\left(W^{2l_r+1}(z_e,z_r) \in \hat{\Omega} | W^1(z_e,z_r) = w^1\right),$$

and $\nu(\Omega) = \nu(\hat{\Omega})$. So if we can prove

$$\mathbb{P}\left(W^{2l_r+1}(z_e,z_r) \in \hat{\Omega} | W^1(z_e,z_r) = w^1\right) \ge \nu(\hat{\Omega}),$$

we can show

$$\mathbb{P}\left(W^{2l_r+1}(z_e,z_r) \in \Omega | W^1(z_e,z_r) = w^1\right) \ge \nu(\Omega).$$

Consider the following demand pattern of length $2l_r$.

$$D^t \le \frac{z_r - z_e}{l_r+1},\ \forall 1 \le t \le l_r + l_e,$$
$$D^{2l_r-k} \in \hat{\Omega}_k,\ \forall k = l-1,\dots,0.$$

This demand pattern happens with probability $\gamma(z_e,z_r)^{l_r+l_e}\cdot\prod_{k=0}^{l-1}\mathbb{P}\left(D\in\hat{\Omega}_k\right)$.

As $\bar{w}^1\cdot\mathbf{1}\le z_r$, we have

$$q_e^1=(z_e-IP_e^1-q_r^{1-l})^+,$$
$$q_r^1=z_r-(IP_e^1+q_r^{1-l}+\ldots+q_r^0+q_e^1),$$

for the first period. Also, from Veeraraghavan and Scheller-Wolf (2008) Eq. (4), we know $q_e^{t+1}+q_r^{t+1}=d^t$, $\forall t\ge 1$. So we have $q_r^{t+1}\le d^t$, $\forall t\ge 1$.

Thus, we have for $l_r+1\le t\le 2l_r$,

$$\begin{aligned}d^t\ge q_r^{t+1}&=z_r-(IP_r^{t+1}+q_e^{t+1})\\&=z_r-IP_r^{t+1}-(z_e-IP_e^{t+1}-q^{t+1-l})^+\\&=z_r-IP_e^{t+1}-q_r^{t-l+2}-\ldots-q_r^t-(z_e-IP_e^{t+1}-q^{t+1-l})^+\\&=z_r-\max(z_e,IP_e^{t+1}+q_r^{t+1-l})-(q_r^{t-l+2}+\ldots+q_r^t).\end{aligned}$$

Therefore

$$\begin{aligned}\max(z_e,IP_e^{t+1}+q_r^{t+1-l})&\ge z_r-d^t-(q_r^{t-l+2}+\ldots+q_r^t)\\&\ge z_r-\frac{l(z_r-z_e)}{l_r+1}>z_e,\end{aligned}$$

because $\dfrac{l}{l_r+1}<1$.

So $\max(z_e,IP_e^{t+1}+q_r^{t+1-l})=IP_e^{t+1}+q_r^{t+1-l}$. Thus, $q_e^{t+1}=0$ and $q_r^{t+1}=d^t$, $\forall l_r+1\le t\le 2l_r$.

Thus, we have

$$(q_r^{2l_r},q_r^{2l_r-1},\ldots,q_r^{l_r+2})=(d^{2l_r-1},d^{2l_r-2},\ldots,d^{l_r+1}).\tag{30}$$

Also,

$$q_r^{2l_r+1}=z_r-IP_e^{2l_r+1}-q_r^{2l_r}-\ldots-q_r^{2l_r+1-l},$$

and therefore,

$$z_r-(IP_e^{2l_r+1}+q_r^{2l_r+1-l})-\sum_{k=2l_r+2-l}^{2l_r}q_r^k=d^{2l_r}.$$

Thus, after the demand pattern, $W^{2l_r+1}(z_e,z_r)=(q_r^{2l_r},q_r^{2l_r-1},\ldots,q_r^{l_r+l_e+2},IP_e^{2l_r+1}+q_r^{2l_r+1-l})\in\hat{\Omega}$ as $q_r^{2l_r}=d^{2l_r-1}\in\hat{\Omega}_1,\ldots,q_r^{l_r+l_e+2}=d^{l_r+l_e+1}\in\hat{\Omega}_{l-1}$, $z_r-(IP_e^{2l_r+1}+q_r^{2l_r+1-l})-\sum_{k=2l_r+2-l}^{2l_r}q_r^k=d^{2l_r}\in\hat{\Omega}_0$. So

$$\mathbb{P}\left(W^{2l_r+1}(z_e,z_r)\in\hat{\Omega}|W^1(z_e,z_r)=w^1\right)\ge\gamma(z_e,z_r)^{l_r+l_e}\cdot\prod_{k=0}^{l-1}\mathbb{P}\left(D\in\hat{\Omega}_k\right)=\nu(\hat{\Omega}).$$

And therefore $\bar{U}$ is a small set with respect to ν and $t^* = 2l_r + 1$. Hence, for any $t \geq 2l_r + 1$, any initial vector $w^1 \in \mathbb{R}_+^{l-1} \times \mathbb{R}$ satisfying $\bar{w}^1 \cdot \mathbf{1}^l \leq z_r$,

$$\delta^{t+1}\left(z_e, z_r, w^1\right) \leq \left(1 - \gamma(z_e, z_r)^{2l_r}\right)^{t/2l_r}.$$

Q.E.D.

B.2. Case 2: Initial regular inventory position exceeds z_r

The proof for this section is similar to the proof of Theorem 3 in Huh et al. (2009) Case 2.

Denote $F(\cdot)$ as the distribution function of D and $\mu = \mathbb{E}[D]$. Below is the Lemma 5 in Huh et al. (2009):

LEMMA 15 (LEMMA 5 IN HUH ET AL. (2009)). *For any $\eta \in \mathbb{R}$ and $t \geq 1$,*

$$\mathbb{P}\left(\sum_{\ell=1}^{t} D^\ell \leq \eta\right) \leq \begin{cases} F(\eta)^t, & \text{if } D \text{ has an infinite support,} \\ e^{4\eta/\bar{D}} \cdot e^{-2t\mu^2/\bar{D}^2}, & \text{if } D \leq \bar{D} \text{ with probability one.} \end{cases}$$

Also, similar to Lemma 6 in Huh et al. (2009), we have the following lemma:

LEMMA 16. *Consider dual-index policy with base stock levels (z_e, z_r). For any regular starting inventory position $w^1 \in \mathbb{R}_+^{l-1} \times \mathbb{R}$ and $t \geq 2l_r$, we have:*

$$\begin{aligned} &\mathbb{P}\left(\bar{W}^t(z_e, z_r) \cdot \mathbf{1}^l > z_r | \bar{W}^1(z_e, z_r) = \bar{w}^1\right) \\ &\leq \begin{cases} F(\bar{w}^1 \cdot \mathbf{1}^l - z_r)^{t-l_r}, & \text{if } D \text{ has infinite support,} \\ e^{4(\bar{w}^1 \cdot \mathbf{1}^l - z_r)/\bar{D}} \cdot e^{-2\mu^2(t-l_r)/\bar{D}^2}, & \text{if } D \leq \bar{D} \text{ with probability one.} \end{cases} \end{aligned}$$

Proof of Lemma 16. According to the base-stock policy in the dual-index policy for the regular inventory position, we have

$$\max\left\{\bar{W}^t(z_e, z_r) \cdot \mathbf{1}^l, z_r\right\} \leq \max\left\{\bar{W}^1(z_e, z_r) \cdot \mathbf{1}^l, z_r\right\}, \forall t \geq 1.$$

So

$$\begin{aligned} &\mathbb{P}\left(D^1 + D^2 + \ldots + D^{t-l_r} < \bar{w}^1 \cdot \mathbf{1}^l - z_r\right) \\ =&\mathbb{P}\left(D^{l_r} + D^{l_r+1} + \ldots + D^{t-1} < \bar{w}^1 \cdot \mathbf{1}^l - z_r\right) \\ \geq&\mathbb{P}\left(D^{l_r} + D^{l_r+1} + \ldots + D^{t-1} < \bar{w}^{l_r} \cdot \mathbf{1}^l - z_r\right). \end{aligned}$$

So $\forall t \geq 2l_r$, we claim that $\bar{W}^t(z_e, z_r) \cdot \mathbf{1}^l > z_r$ if and only if $\bar{W}^{l_r}(z_e, z_r) \cdot \mathbf{1}^l - (D^{l_r} + D^{l_r+1} + \ldots + D^{t-1}) > z_r$.

If $\bar{W}^t(z_e, z_r) \cdot \mathbf{1}^l > z_r$, then according to the dual-index policy, we have $q_r^k = 0$, $\forall k \leq t$. So we have $\bar{W}^t(z_e, z_r) \cdot \mathbf{1}^l = \max(z_e, IP_e^t)$.

Because $z_e \leq z_r$, we have $\bar{W}^t(z_e, z_r) \cdot \mathbf{1}^l = IP_e^t > z_r \geq z_e$. Also, $\bar{W}^{l_r}(z_e, z_r) \cdot \mathbf{1}^l = IP_e^{l_r} > z_r \geq z_e$. So $q_e^k = 0$, $\forall l_r \leq k \leq t$.

So

$$\bar{W}^{l_r}(z_e, z_r) \cdot \mathbf{1}^l - (D^{l_r} + D^{l_r+1} + \ldots + D^{t-1})$$

$$
\begin{aligned}
&= IP_e^{l_r} - (D^{l_r} + D^{l_r+1} + \ldots + D^{t-1}) \\
&= IP_e^t > z_r.
\end{aligned}
$$

If $\bar{W}^{l_r}(z_e, z_r) \cdot \mathbf{1}^l - (D^{l_r} + D^{l_r+1} + \ldots + D^{t-1}) > z_r$, then $\bar{W}^t(z_e, z_r) \cdot \mathbf{1}^l > z_r$.

Hence, $\bar{W}^t(z_e, z_r) \cdot \mathbf{1}^l > z_r$ if and only if $\bar{W}^{l_r}(z_e, z_r) \cdot \mathbf{1}^l - (D^{l_r} + D^{l_r+1} + \ldots + D^{t-1}) > z_r$.

Thus, by Lemma 15, the results of Lemma 16 follow. **Q.E.D.**

Next, we only prove the result when D has infinite support. The proof for the situation when D is bounded is similar.

Let $\mathbb{P}_{\bar{w}^1}$ and $\mathbb{E}_{\bar{w}^1}$ be the probability and expectation conditioned on the event that $\bar{W}^1 \cdot \mathbf{1}^l = \bar{w}^1$. Then for any measurable set $\Omega \subseteq \mathbb{R}_+^{l-1} \times \mathbb{R}$,

$$
\begin{aligned}
&\mathbb{P}_{\bar{w}^1}[W^{t+1}(z_e, z_r) \in \Omega] \\
=&\mathbb{E}_{\bar{w}^1}\left[\mathbb{P}_{\bar{w}^1}[W^{t+1}(z_e, z_r) \in \Omega \mid W^{\lceil \frac{t}{2} \rceil}(z_e, z_r)]\right] \\
=&\mathbb{E}_{\bar{w}^1}\left[\mathbb{1}(W^{\lceil \frac{t}{2} \rceil}(z_e, z_r) \cdot \mathbf{1}^l \le z_r) \cdot \mathbb{P}\left(W^{t+1}(z_e, z_r) \in \Omega | W^{\lceil \frac{t}{2} \rceil}(z_e, z_r)\right)\right] \\
&+ \mathbb{E}_{\bar{w}^1}\left[\mathbb{1}(\bar{W}^{\lceil \frac{t}{2} \rceil}(z_e, z_r) \cdot \mathbf{1}^l > z_r) \cdot \mathbb{P}\left(W^{t+1}(z_e, z_r) \in \Omega | \bar{W}^{\lceil \frac{t}{2} \rceil}(z_e, z_r)\right)\right)],
\end{aligned}
$$

where the last equality is from Markov property. Then we have

$$
\begin{aligned}
A :=& \mathbb{P}_{\bar{w}^1}[W^{t+1}(z_e, z_r) \in \Omega] - \mathbb{P}\left(W^{\infty}(z_e, z_r) \in \Omega\right) \\
=& \mathbb{E}_{\bar{w}^1}\left[\mathbb{1}(\bar{W}^{\lceil \frac{t}{2} \rceil}(z_e, z_r) \cdot \mathbf{1}^l \le z_r) \cdot \phi(\Omega)\right] + \mathbb{E}_{\bar{w}^1}\left[\mathbb{1}(\bar{W}^{\lceil \frac{t}{2} \rceil}(z_e, z_r) \cdot \mathbf{1}^l > z_r) \cdot \phi(\Omega)\right],
\end{aligned}
$$

where $\phi(\Omega) = \mathbb{P}\left(W^{t+1}(z_e, z_r) \in \Omega | W^{\lceil \frac{t}{2} \rceil}(z_e, z_r)\right) - \mathbb{P}\left(W^{\infty}(z_e, z_r) \in \Omega\right)$.

Also, we have almost surely

$$
|\phi(\Omega)| \le \delta_{t-\lceil \frac{t}{2} \rceil + 2}(z_e, z_r, W^{\lceil \frac{t}{2} \rceil}(z_e, z_r)),
$$

and $\phi(\Omega) \le 1$, so

$$
\begin{aligned}
|A| \le& \mathbb{E}_{\bar{w}^1}\left[\mathbb{1}(\bar{W}^{\lceil \frac{t}{2} \rceil}(z_e, z_r) \cdot \mathbf{1}^l \le z_r) \cdot \delta_{t-\lceil \frac{t}{2} \rceil + 2}(z_e, z_r, W^{\lceil \frac{t}{2} \rceil}(z_e, z_r))\right] \\
&+ \mathbb{P}_{\bar{w}^1}[\bar{W}^{\lceil \frac{t}{2} \rceil}(z_e, z_r) \cdot \mathbf{1}^l > z_r].
\end{aligned}
$$

By Lemma 14, the first term

$$
\begin{aligned}
&\mathbb{E}_{\bar{w}^1}\left[\mathbb{1}(\bar{W}^{\lceil \frac{t}{2} \rceil}(z_e, z_r) \cdot \mathbf{1}^l \le z_r) \cdot \delta_{t-\lceil \frac{t}{2} \rceil + 2}(z_e, z_r, W^{\lceil \frac{t}{2} \rceil}(z_e, z_r))\right] \\
\le& \mathbb{P}_{\bar{w}^1}[\bar{W}^{\lceil \frac{t}{2} \rceil}(z_e, z_r) \cdot \mathbf{1}^l \le z_r] \cdot \left(1 - \gamma(z_e, z_r)^{2l_r}\right)^{(t-\lceil \frac{t}{2} \rceil + 1)/2l_r} \\
\le& \left(1 - \gamma(z_e, z_r)^{2l_r}\right)^{(t-\lceil \frac{t}{2} \rceil + 1)/2l_r} \\
\le& \left(1 - \gamma(z_e, z_r)^{2l_r}\right)^{t/4l_r}.
\end{aligned}
$$

By Lemma 16, the second term

$$
\mathbb{P}_{\bar{w}^1}[\bar{W}^{\lceil \frac{t}{2} \rceil}(z_e, z_r) \cdot \mathbf{1}^l > z_r] \le F(\bar{w}^1 \cdot \mathbf{1}^l - z_r)^{\lceil \frac{t}{2} \rceil - l_r}
$$

$$\leq F(\bar{w}^1 \cdot \mathbf{1}^l - z_r)^{\frac{t}{2}-l_r}.$$

Therefore, we obtain the bound for $\delta^{t+1}(z_e, z_r, w^1)$ as stated in Theorem 1.

Appendix C: Proof of Lemmas

C.1. Proof of Lemma 8

Because

$$L^{n-1} = \sum_{i=1}^{n-1} \left\lceil \frac{2^i}{\log T} \right\rceil \geq \sum_{i=1}^{n-1} \frac{2^i}{\log T} = \frac{2^n - 2}{\log T},$$

we have $\alpha^n \leq \frac{3}{2}\sqrt{3T_0}\dfrac{\log T}{\sqrt{2^n-2}}$. Therefore,

$$\begin{aligned}
\sum_{n=1}^{N} B^n \alpha^n \leq & \sum_{n=1}^{N} \left\lceil \frac{2^n}{\log T} \right\rceil \frac{3}{2}\sqrt{3T_0}\frac{\log T}{\sqrt{2^n-2}} \\
\leq & \frac{3}{2}\sqrt{3T_0}\sum_{n=1}^{N}\left(\frac{2^n}{\sqrt{2^n-2}} + \sqrt{\frac{\log T}{2^n-2}}\right) \\
= & O(\sqrt{T\log T}),
\end{aligned}$$

where the last line is because $N \leq \log_2(T\log T + 2) - 1$.

C.2. Proof of Lemma 12

For any epoch $n \in [N]$ and any arm $j \in \mathcal{A}^n$, as $\{W_j^t\}_{t=1}^{L^n}$ is the Markov chain of the states of the system following the dual-index policy $\left(\Delta_j, z_{ej}^n\right)$, let

$$\begin{aligned}
g^n\left(\hat{W}_j^1, \ldots, \hat{W}_j^{L^n}\right) &= \frac{1}{L^n}\sum_{t=1}^{L^n} \hat{C}^t\left(\Delta_j, z_{e_j}^n\right) \\
&= \frac{1}{L^n}\sum_{t=1}^{L^n} c_e \hat{q}_{ej}^t + c_r \hat{q}_{r_j}^t + h\left(\hat{I}_j^{t+1}\right)^+ + b\left(\hat{I}_j^{t+1}\right)^-
\end{aligned}$$

Also, notice that condition (8) holds with $\iota_i = \frac{\bar{C}}{L^n}$, $\forall i \in [L^n]$ as $C^t(\Delta, z_e) \leq \bar{C}$ with probability 1 for any Δ and z_e. Then Lemma 12 holds according to Lemma 6 with Markov chain being $\{W_j^t\}_{t=1}^{L^n}$ and the function f being g^n for any $n \in [N]$ and $j \in [J]$.

C.3. Proof of Lemma 13

For notational simplicity, let $\mathbb{E}[C^*(\Delta)] := \mathbb{E}[C^\infty(\Delta, z_e^*(\Delta))]$. For Δ_1 and Δ_2, without loss of generality, we assume $\mathbb{E}[C^*(\Delta_1)] \geq \mathbb{E}[C^*(\Delta_2)]$. Let $z_1 = \arg\min_z \mathbb{E}[C^\infty(\Delta_1, z_e)]$ and $z_2 = \arg\min_z \mathbb{E}[C^\infty(\Delta_2, z_e)]$.

$$\begin{aligned}
&\mathbb{E}[C^*(\Delta_1)] - \mathbb{E}[C^*(\Delta_2)] \\
=&\mathbb{E}[C^\infty(\Delta_1, z_1)] - \mathbb{E}[C^\infty(\Delta_2, z_2)] \\
\leq&\mathbb{E}[C^\infty(\Delta_1, z_2)] - \mathbb{E}[C^\infty(\Delta_2, z_2)] \\
=&(c_e + c_r)\mathbb{E}[O^\infty(\Delta_1) - O^\infty(\Delta_2)] + h\mathbb{E}\left[(z_2 + O^\infty(\Delta_1) - D_{l_e})^+ - (z_2 + O^\infty(\Delta_2) - D_{l_e})^+\right] \\
&+ b\mathbb{E}\left[(z_2 + O^\infty(\Delta_1) - D_{l_e})^- - (z_2 + O^\infty(\Delta_2) - D_{l_e})^-\right].
\end{aligned}$$

We would like to offer an upper bound for the term $\mathbb{E}\left|O^{\infty}(\Delta_1)-O^{\infty}(\Delta_2)\right|$. First, we show that $\mathbb{E}\left[O^{\infty}(\Delta)\right]$ is non-decreasing in Δ by contradiction.

Suppose that $\Delta_1 \geq \Delta_2$ with $\mathbb{E}\left[O^{\infty}(\Delta_1)\right] < \mathbb{E}\left[O^{\infty}(\Delta_2)\right]$. From Equation (8) in Veeraraghavan and Scheller-Wolf (2008) which is $O^{\infty}(\Delta) = \Delta - (q_r^t + q_r^{t-1} + \ldots + q_r^{t-l+1})$, we have $\mathbb{E}\left[O^t(\Delta)\right] = \Delta - l\mathbb{E}\left[q_r^{\infty}\right]$. Consider the following two cases:

1. If $\mathbb{E}\left[O^{\infty}(\Delta_1)\right] = 0$, which means $\mathbb{E}\left[q_e^{\infty}\right] \geq 0$ and thus $\mathbb{E}\left[q_r^{\infty}\right] \leq \mathbb{E}\left[D\right]$.

$$\begin{aligned}&\mathbb{E}\left[O^{\infty}(\Delta_1)\right] - \mathbb{E}\left[O^{\infty}(\Delta_2)\right]\\ =&\Delta_1 - l\mathbb{E}\left[q_r^{\infty}(\Delta_1)\right] - \Delta_2 + l\mathbb{E}\left[D\right].\end{aligned}$$

   As $\mathbb{E}\left[q_r^{\infty}(\Delta_1)\right] \leq l\mathbb{E}\left[D\right]$ and $\Delta_1 \geq \Delta_2$, we have

$$\mathbb{E}\left[O^{\infty}(\Delta_1)\right] - \mathbb{E}\left[O^{\infty}(\Delta_2)\right] \geq 0,$$

   which is a contradiction.

2. If $\mathbb{E}\left[O^{\infty}(\Delta_1)\right] > 0$, which means $\mathbb{E}\left[q_r^{\infty}\right] = \mathbb{E}\left[D\right]$. Then we have

$$\begin{aligned}&\mathbb{E}\left[O^{\infty}(\Delta_1)\right] - \mathbb{E}\left[O^{\infty}(\Delta_2)\right]\\ =&\Delta_1 - l\mathbb{E}\left[D\right] - \Delta_2 + l\mathbb{E}\left[D\right] \geq 0,\end{aligned}$$

   which is a contradiction.

Therefore, we have $\mathbb{E}\left[O^{\infty}(\Delta)\right]$ is non-decreasing in Δ. So consider $\Delta_1 \geq \Delta_2$ without loss of generality. Then we have $\mathbb{E}\left[O^{\infty}(\Delta_1)\right] \geq \mathbb{E}\left[O^{\infty}(\Delta_2)\right]$.

1. If $\mathbb{E}\left[O^{\infty}(\Delta_1)\right] > 0$ and $\mathbb{E}\left[O^{\infty}(\Delta_2)\right] > 0$, then

$$\mathbb{E}\left[O^{\infty}(\Delta_1)\right] - \mathbb{E}\left[O^{\infty}(\Delta_2)\right] = \Delta_1 - \Delta_2.$$

2. If $\mathbb{E}\left[O^{\infty}(\Delta_1)\right] > 0$ and $\mathbb{E}\left[O^{\infty}(\Delta_2)\right] = 0$, then

$$\mathbb{E}\left[O^{\infty}(\Delta_1)\right] - \mathbb{E}\left[O^{\infty}(\Delta_2)\right] = \Delta_1 - \Delta_2 - l\mathbb{E}\left[D\right] + l\mathbb{E}\left[q_r^{\infty}(\Delta_2)\right] \leq \Delta_1 - \Delta_2.$$

3. If $\mathbb{E}\left[O^{\infty}(\Delta_1)\right] = 0$ and $\mathbb{E}\left[O^{\infty}(\Delta_2)\right] = 0$, then

$$\mathbb{E}\left[O^{\infty}(\Delta_1)\right] - \mathbb{E}\left[O^{\infty}(\Delta_2)\right] = 0.$$

In sum, $\mathbb{E}\left[\left|O^{\infty}(\Delta_1) - O^{\infty}(\Delta_2)\right|\right] \leq \left|\Delta_1 - \Delta_2\right|$.

Then, because $(x-a)^+ - (y-a)^+ \leq |x-y|$ and $(x-a)^- - (y-a)^- \leq |x-y|$, we have

$$\mathbb{E}\left[C^*(\Delta_1)\right] - \mathbb{E}\left[C^*(\Delta_2)\right] \leq (c_e + c_r + h + b)\mathbb{E}\left[\left|O^{\infty}(\Delta_1) - O^{\infty}(\Delta_2)\right|\right] \leq (c_e + c_r + h + b)\left|\Delta_1 - \Delta_2\right|,$$

which suggests that $\mathbb{E}\left[C^*(\Delta)\right]$ is Lipschitz in Δ.

Appendix D: Settings with Nonstationary Demand

The i.i.d. demand assumption is predominant in the dual sourcing literature (Allon and Van Mieghem 2010, Sheopuri et al. 2010). The dual-index policy was, therefore, developed for stationary demand (Veeraraghavan and Scheller-Wolf 2008), and its performance has never been explored under *non-stationary* demand.

Restart Learning Algorithm. When demand is non-i.i.d., the performance guarantee of our (Δ, z_e) algorithm may not hold (see Figure 10) if our learning algorithm is naively implemented in this nonstationary environment. Thus, we need to come up with an alternative strategy. Borrowing the "restart" idea from Besbes et al. (2015), we re-design our algorithm by restarting the procedure for every τ_T periods. The details of our modified algorithm are in Algorithm 2. Regarding the choice of τ_T, in Besbes et al. (2015), $\tau_T = \left\lceil (T/V_T)^{2/3} \right\rceil$ where V_T is a known variation budget. In our problem, V_T is defined as the upper bound of $\sum_{t=2}^{T} \left\| C_t^{\infty} - C_{t-1}^{\infty} \right\| = \sum_{t=2}^{T} \sup_{(z_e,z_r)\in[0,\bar{Z}]\times[0,\bar{Z}]} |C_t^{\infty}(z_e,z_r) - C_{t-1}^{\infty}(z_e,z_r)|$ where $C_t^{\infty}(\cdot)$ is the stationary per-period cost when demand follows the same distribution as D^t. Note that our algorithm is different from the OGD studied in Besbes et al. (2015) and the f function is not necessarily convex, the tuning of the restarting interval will vary. For experimental purposes, we follow the choice of $\tau_T = \left\lceil (T/V_T)^{2/3} \right\rceil$ based on the intuition that the larger the variation budget is, the smaller the restart interval should be. Since the performance measure is the relative regret, the cost parameters can be scaled and so is the variation budget. We assume that the firm knows the variation budget $V_T = 1$ after rescaling the cost parameters.

Convergence Rate. Here we briefly analyze the performance guarantee of the proposed Algorithm 2 denoted as π'. Since the optimal inventory replenishment policy is complex and state-dependent even under stationary demand, we still choose the full-information optimal dual-index policies as the benchmark under non-stationary demand. Specifically, for any algorithm **ALG**, we define the performance metric under non-stationary demand as

$$\mathcal{R}_T^{\mathbf{ALG}} = \mathbb{E}\left[\sum_{t=1}^{T} C_{\mathbf{ALG}}^t - \sum_{t=1}^{T} C_t^{\infty}\left(z_e^{t*}, z_r^{t*}\right) \right],$$

where $C_{\mathbf{ALG}}^t$ is the cost in period t by running algorithm **ALG**. We define $C_t^{\infty}(z_e, z_r)$ as the stationary cost variable under the dual-index policy with dual indices (z_e, z_r) under the demand with the same distribution as D^t and $(z_e^{t*}, z_r^{t*}) := \arg\max_{(z_e,z_r)} C_t^{\infty}(z_e, z_r)$.

We conjecture that $\mathcal{R}_T^{\pi'} = \tilde{O}\left(T^{\frac{2}{3}} V_T^{\frac{1}{3}}\right)$, but we leave the rigorous proof to future work. Here, we only provide some intuitions and technical results, which would help build the foundation of a rigorous proof.

It is noteworthy that the establishment of Lemma 1 and Lemma 2 does not rely on the stationarity of the demand distribution, and both lemmas still hold (with an additional assumption restricting the cumulative demand from being excessively small for Lemma 2). In particular, $(W^t(z_e, z_r), t \geq 1)$ in the dual-sourcing system following a dual-index policy with parameters (z_e, z_r) under non-stationary demand still forms a (not necessarily homogeneous) Markov chain. Moreover, two processes driven by the same demand will couple after $O(\log T)^2$ periods with high probability.

As for another key result Lemma 6, which guarantees the performance of the empirical estimation framework, we here define the mixing time for Markov chains without assuming time homogeneity. We let $\mathcal{L}(X_{i+t} \mid X_i = x)$ be the conditional distribution of X_{i+t} given $X_i = x$.

$$\bar{d}(t) := \max_{1 \le i \le N-t} \sup_{x,y \in \Omega_i} d_{\mathrm{TV}}\left(\mathcal{L}(X_{i+t} \mid X_i = x), \mathcal{L}(X_{i+t} \mid X_i = y)\right),$$
$$\tau(\epsilon) := \min\{t \in \mathbb{N} : \bar{d}(t) \le \epsilon\}.$$

The following generalized result of Lemma 6 exists for not necessarily homogeneous Markov chains.

LEMMA 17 (COROLLARY 2.10 IN PAULIN (2015)). *Let $X := (X_1, \ldots, X_N)$ be a (not necessarily time-homogeneous) Markov chain, taking values in a Polish state space $\Lambda = \Lambda_1 \times \ldots \times \Lambda_N$, with mixing time $\tau(\epsilon)$(for $0 \le \epsilon \le 1$). Let $\tau_{\min} := \inf_{0 \le \epsilon < 1} \tau(\epsilon) \cdot \left(\frac{2-\epsilon}{1-\epsilon}\right)^2$. Suppose that $f : \Lambda \to \mathbb{R}$ satisfies $f(x) - f(y) \le \sum_{i=1}^n c_i \mathbb{1}[x_i \ne y_i]$ for every $x, y \in \Lambda$. Then for any $t \ge 0$,*

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \ge t) \le 2\exp\left(\frac{-2t^2}{\|c\|^2 \tau_{\min}}\right).$$

Thus, we have a concentration inequality established for the sample average up to time t of some function with respect to the Markov chains with non-stationary transition kernels, compared with the true mean up to time t of the function under this nonstationary environment. With proper assumptions on the degree of changes in the underlying demand distribution, $\tau_{\min}$ would be of constant order. Consequently, following similar proof as in the case of a stationary environment, we can offer an upper bound for the difference between the cost incurred by the original (Δ, z_e) Algorithm (denoted as π) and the *static* optimal dual-index policy, i.e., $\mathscr{R}^{\pi}_{\tau_T} := \mathbb{E}\left[\sum_{t=1}^{\tau_T} C^t_{\pi} - \tau_T C^{\infty}(z^*_e, z^*_r)\right]$ where $(z^*_e, z^*_r) := \arg\max_{(z_e, z_r)} \mathbb{E}\left[\sum_{t=1}^{\tau_T} C^{\infty}_t(z_e, z_r)\right]$. Since all key results hold in the same order as in stationary cases, we speculate that $\mathscr{R}^{\pi}_{\tau_T} = \tilde{O}(\sqrt{\tau_T})$.

Combined with the following proposition, we can provide the upper bound for the dynamic regret.

PROPOSITION 8. (PROPOSITION 2 IN BESBES ET AL. (2015)) *Let π' be the policy defined by the restarting procedure that uses π as a subroutine with batch size τ_T. Then, for any $T \geqslant 1$,*

$$\mathcal{R}^{\pi'}_T \leqslant \left\lceil \frac{T}{\tau_T} \right\rceil \cdot \mathscr{R}^{\pi}_{\tau_T} + 2\tau_T V_T.$$

Because the establishment of Proposition 8 does not require the convexity property of the objective function, adopting the restarting procedure with batch size $\tau_T = \left\lceil (T/V_T)^{2/3} \right\rceil$ will incur a total regret $\mathcal{R}^{\pi'}_T = \tilde{O}\left(T^{\frac{2}{3}} V_T^{\frac{1}{3}}\right)$, matching the information-theoretic lower bound established in Besbes et al. (2015) up to logarithmic factors. If we consider the instance below where $V_T = O(1)$ and $\tau_T = \left\lceil (T/V_T)^{2/3} \right\rceil = 150$, the total regret is of order $O(T^{\frac{2}{3}} \log T)$ as shown in Figure 8.

Non-IID Demand Instance. Consider the following test instance where the time horizon is five consecutive years with $T = 1825$. The lead times are set to be $l_e = 2, l_r = 4$. The demand in period t is set to follow the truncated normal distribution $\mathcal{N}(\mu_n, \sigma_n^2)$ where $n = \lceil t/365 \rceil$ as shown in Table 6.

Table 6 Demand Settings

Year	1	2	3	4	5
Distribution	$\mathcal{N}(30,6)$	$\mathcal{N}(40,7)$	$\mathcal{N}(50,10)$	$\mathcal{N}(60,12)$	$\mathcal{N}(70,14)$
Truncated at	$[0,60]$	$[10,70]$	$[20,80]$	$[30,90]$	$[40,100]$

Computational Performance. We emphasize that the firm does not know the evolution of the demand distribution nor the distributions themselves when running the learning algorithm. The relative regret is defined as

$$\textbf{Relative Regret} := \frac{\sum_{t=1}^{T} C^t(z_e^{t\pi}, z_r^{t\pi}) - \sum_{t=1}^{T} C^t(z_e^{t*}, z_r^{t*})}{\sum_{t=1}^{T} C^t(z_e^{t*}, z_r^{t*})},$$

where (z_e^{t*}, z_r^{t*}) are the optimal order-up-to levels for the dual-sourcing system with demand following $\mathcal{N}(\mu_{\lceil t/365 \rceil}, \sigma^2_{\lceil t/365 \rceil})$. Figure 8 shows the relative regret averaged over the instances with (b,h) pairs taking values $(\{5,10\} \times \{1,4\})$ as in Table 1. For each instance, we run 1000 times and take the average of the relative regret. Also, Figure 9 shows the performance of the restart (Δ, z_e) algorithm when the restart point is tuned to be the change point. For comparison, Figure 10 shows the performance of the (Δ, z_e) algorithm without restart. It is evident that the restarting procedure is necessary under non-stationary demand, and the modified restart algorithm works well when the restarting point is close to the change point of demand.

Unknown Variation Budget. Algorithm 2 requires prior knowledge of the total variation budget V_T to determine the length of restarting epoch τ_T. When there is no information on the degree of non-stationarity, one can schedule multiple instances of the base algorithm with different durations in a carefully-designed randomized scheme and restart based on the real-time detection result of the change of the environment as introduced in Wei and Luo (2021). The design and analysis of this framework for the dual-index policy in dual-sourcing systems are left to future work.

Algorithm 2 The "restart" (Δ, z_e) learning algorithm for the dual-index policy

for $k = 1, \ldots, \lceil T/\tau_T \rceil$: **do** ▷ **Restart**

Let $T_0 = (\min\{\tau_T, T-(k-1)\tau_T\})$ and $N = \min\left\{n : \sum_{i=1}^{n} \lceil \frac{2^i}{\log T_0} \rceil \geq T_0\right\}$ the number of epochs.

Let $J = \#$ discrete Δ's and $B^i = \lceil \frac{2^i}{\log T_0} \rceil$ be the i-th epoch length.

Let $L^n = \sum_{i=1}^{n} B^i$, $\forall n \in [N-1]$ with $L^0 = 0, L^N = T_0$. ▷ **Parameters**

Initialize the active set $\mathcal{A}^1 = \{1, \ldots, J\}$, $\mathcal{D}^0 = \varnothing$. ▷ **Initialization**

For $j \in \mathcal{A}^1$, define $\Delta_j = \underline{\Delta} + \frac{j}{J}\left|\bar{Z} - \underline{\Delta}\right|$ and assign $z_{ej}^1 \in [0, \bar{Z} - \Delta_j]$ arbitrarily.

for $n = 1, 2, \ldots, N$ **do** ▷ **Outer Loop**

Randomly select $j^n \in \mathcal{A}^n$. Let demand set $\mathcal{D}^n = \mathcal{D}^{n-1}$.

for $t = (k-1)\tau_T + L^{n-1} + 1, \ldots, (k-1)\tau_T + L^n$: **do**

Apply the dual-index policy $(z_e^t, z_r^t) = (z_{ej^n}^n, z_{ej^n}^n + \Delta_{j^n})$.

Append the realized demand d^t into $\mathcal{D}^n$.

$$
\begin{aligned}
&q_e^t = (z_{ej^n}^n - IP_e^t - q_r^{t-l})^+, \quad q_r^t = (z_{ej^n}^n + \Delta_{j^n} - IP_r^t - q_e^t)^+,\\
&IP_e^{t+1} = IP_e^t + q_e^t - d^t + q_r^{t-l}, \quad IP_r^{t+1} = IP_r^t + q_e^t + q_r^t - d^t,\\
&o^t = (IP_e^t + q_r^{t-l} - z_{ej^n}^n)^+, \quad I^{t+1} = I^t + q_e^{t-l_e} + q_r^{t-l_r} - d^t.
\end{aligned}
$$

end for

for $j \in \mathcal{A}^n$ **do** ▷ **Inner Loop**

Simulate the policy $(z_{ej}^n, z_{ej}^n + \Delta_j)$ for $\min\{L^n, T_0\}$ periods using $\mathcal{D}^n$ and denote the state variables of this simulation by $\hat{W}_j^t := (\hat{q}_{rj}^{t-1}, \ldots, \hat{q}_{rj}^{t-l+1}, \widehat{IP}_{ej}^t + \hat{q}_{rj}^{t-l}) \in \mathbb{R}_+^{l-1} \times \mathbb{R}$, $t = (k-1)\tau_T + 1, \ldots, (k-1)\tau_T + L^n$.

Obtain the estimated average period cost:

$$
\hat{G}_j^n = \frac{1}{L^n} \sum_{t \in [L^n]} c_e \hat{q}_{ej}^t + c_r \hat{q}_{rj}^t + h(\hat{I}_j^{t+1})^+ + b(\hat{I}_j^{t+1})^-.
$$

Let $\mathcal{X}_j^n = \left\{d_{l_e}^t - \hat{o}_j^t, t \in [(k-1)\tau_T + 1, (k-1)\tau_T + L^n - l_e]\right\}$.

Let $\hat{F}_{\Delta_j}^n(\cdot)$ be the empirical CDF of $X_j^t = D_{l_e}^t - \hat{O}_j^t(\Delta_j)$ with data sample $\mathcal{X}_j^n$.

Update $z_{ej}^{n+1} = \hat{F}_{\Delta_j}^{n^{-1}}(\frac{b}{b+h})$. ▷ **Inner Layer Optimization**

end for

Update and prune the active set ▷ **Outer Layer Optimization**

$$
\mathcal{A}^{n+1} = \left\{j \in \mathcal{A}^n : \hat{G}_j^n - \min_{j' \in \mathcal{A}^n} \hat{G}_{j'}^n \leq \varepsilon^n\right\}.
$$

end for

end for

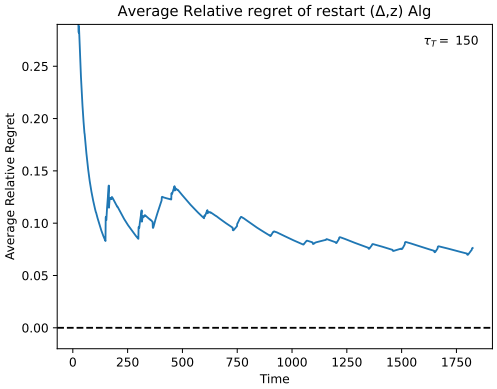

Figure 8 **Restart Interval** $\tau_T = 150$

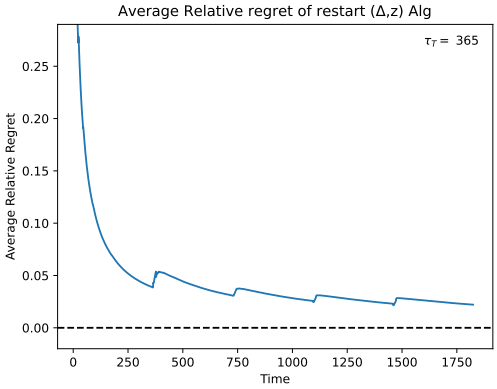

Figure 9 **Restart Interval** $\tau_T = 365$

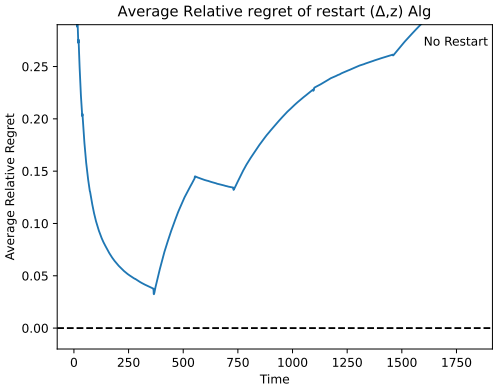

Figure 10 (Δ, z_e) **Algorithm without Restart**