Dynamic Learning with Frequent New Product Launches: A Sequential Multinomial Logit Bandit Problem

Supplementary Material

Junyu Cao Wei Sun

Proposition 4 *The optimal product offering $\mathbf{S}^*$ to the optimization problem (3.1) in each tier is a profit-ordered set. In addition, $\mathbf{S}^*$ is profit-ordered by tier.*

Proof. To show S_1^* is profit-ordered, we will first prove that for any $i \in S_1^*$, $r_i \geq E[R(\mathbf{S}^*)]$. Suppose there exists $i \in S_1^*$ and $r_i < E[R(\mathbf{S}^*)]$, it implies that,

$$
\begin{aligned}
& r_i < E[R(\mathbf{S}^*)] = \frac{\sum_{k\in S_1^*\backslash\{i\}} v_k r_k + r_i v_i + E[R(S_2^*)]}{1+\sum_{k\in S_1^*\backslash\{i\}} v_k + v_i} \\
\Leftrightarrow & \left(1+\sum_{k\in S_1^*\backslash\{i\}} v_k\right) r_i < \sum_{k\in S_1^*\backslash\{i\}} v_k r_k + E[R(S_2^*)] \\
\Leftrightarrow & \left(1+\sum_{k\in S_1^*\backslash\{i\}} v_k\right) r_i v_i + \left(\sum_{k\in S_1^*\backslash\{i\}} v_k r_k + E[R(S_2^*)]\right)\left(1+\sum_{k\in S_1^*\backslash\{i\}} v_k\right) \\
& < v_i \left(\sum_{k\in S_1^*\backslash\{i\}} v_k r_k + E[R(S_2^*)]\right) + \left(\sum_{k\in S_1^*\backslash\{i\}} v_k r_k + E[R(S_2^*)]\right)\left(1+\sum_{k\in S_1^*\backslash\{i\}} v_k\right) \\
\Leftrightarrow & \left(1+\sum_{k\in S_1^*\backslash\{i\}} v_k\right)\left(r_i v_i + \sum_{k\in S_1^*\backslash\{i\}} v_k r_k + E[R(S_2^*)]\right) \\
& < \left(1+\sum_{k\in S_1^*\backslash\{i\}} v_k + v_i\right)\left(\sum_{k\in S_1^*\backslash\{i\}} v_k r_k + E[R(S_2^*)]\right) \\
\Leftrightarrow & \frac{\sum_{k\in S_1^*\backslash\{i\}} v_k r_k + r_i v_i + E[R(S_2^*)]}{1+\sum_{k\in S_1^*\backslash\{i\}} v_k + v_i} < \frac{\sum_{k\in S_1^*\backslash\{i\}} v_k r_k + E[R(S_2^*)]}{1+\sum_{k\in S_1^*\backslash\{i\}} v_k}.
\end{aligned}
$$

The last inequality suggests that removing i from S_1^* will improve the expected profit. It is a contradiction that $\mathbf{S}^*$ is optimal. Similarly we can prove that for any $i \notin S_1^*$, $r_i \leq E[R(\mathbf{S}^*)]$, otherwise adding i to S_1^* will improve the expected profit. The proof follows similarly for S_2^*.

To prove $\mathbf{S}^*$ is profit-ordered by level, notice that the expected profit of the tiered assortment is at least as large as only offering S_2 since $\mathbf{S} = (\emptyset, S_2)$ is also a feasible solution. That is, $E[R(S_2^*)] \leq E[R(\mathbf{S}^*)]$. Therefore, $r_i \leq E[R(S_2^*)] \leq E[R(\mathbf{S}^*)] \leq r_j$ for any $i \notin S_2^*, j \in S_1^*$. Thus, we reach the desired result. ∎

Proposition 5 *Denote the optimal recommendation before and after including a new product with profit r_m to a candidate set as $\mathbf{S}^* = (S_1^*, S_2^*, \cdots, S_W^*)$ and $\hat{\mathbf{S}}^*$, respectively. Define $\mathbf{S}_j^* = (S_j^*, S_{j+1}^*, \cdots, S_W^*)$. The following properties holds.*

a.) $E[R(\hat{\mathbf{S}}_j^*)] \geq E[R(\hat{\mathbf{S}}_{j+1}^*)]$ *for any* $j = 1, \cdots, W-1$.

b.) *If* $E[R(\mathbf{S}_j^*)] < r_m < E[R(\mathbf{S}_{j-1}^*)]$ *for some* j, *then* $m \in \hat{\mathbf{S}}^*$ *but* $m \notin \hat{S}_1^* \cup \hat{S}_2^* \cup \cdots \cup \hat{S}_{j-1}^*$.

c.) *If* $r_m < E[R(S_W^*)]$, *then* $m \notin \hat{\mathbf{S}}^*$.

Proof. First note the relation between $E[R(\hat{\mathbf{S}}_j^*)]$ and $E[R(\hat{\mathbf{S}}_{j+1}^*)]$, where

$$E[R(\hat{\mathbf{S}}_j^*)] = \frac{\sum_{i \in \hat{\mathbf{S}}_j^*} r_i v_i + E[R(\hat{\mathbf{S}}_{j+1}^*)]}{1 + \sum_{i \in \hat{\mathbf{S}}_j^*} v_i}.$$

Therefore, for any j, we have $E[R(\hat{\mathbf{S}}_j^*)] \geq E[R(\hat{\mathbf{S}}_{j+1}^*)]$, otherwise $\mathbf{S} = (\hat{S}_1^*, \cdots, \hat{S}_{j-1}^*, \emptyset, \hat{\mathbf{S}}_{j+1}^*)$ would be a feasible solution whose expected profit is higher than $E[R(\hat{\mathbf{S}}^*)]$. It is a contradiction to the fact that $\hat{\mathbf{S}}^*$ is optimal. This proves a).

Note that $E[R(\hat{\mathbf{S}}_j^*)] \geq E[R(\mathbf{S}_j^*)]$ for any j since $\hat{\mathbf{S}}_j^*$ comes from a larger candidate pool. Following the similar procedure in the proof for Proposition 4, if the new product m is included in tier k where $k \leq j-1$ but $r_m < E[R(\mathbf{S}_{j-1}^*)] \leq E[R(\mathbf{S}_k^*)] \leq E[R(\hat{\mathbf{S}}_k^*)]$, we prove *b*) and *c*) by contradiction. That is, removing product m in tier k will make the expected profit higher. It implies that $m \notin \hat{S}_1^* \cup \hat{S}_2^* \cup \cdots \cup \hat{S}_{j-1}^*$. On the other hand, if m is not included in tier $j, \cdots, W$, then we have $E[R(\mathbf{S}_i^*)] = E[R(\hat{\mathbf{S}}_i^*)]$ for all $i = j, \cdots, W$. Therefore if $r_m > E[R(\mathbf{S}_j^*)] = E[R(\hat{\mathbf{S}}_j^*)] \geq \cdots \geq E[R(\mathbf{S}_W^*)] = E[R(\hat{\mathbf{S}}_W^*)]$, then adding m to any one of the tier $j, \cdots, W$ would increase the expected profit, which is also a contradiction to the fact that $\hat{\mathbf{S}}^*$ is optimal. It concludes the proof for *b*) and *c*). ■

Lemma 7 $\hat{v}_{i,l}^{(k)}$ *are i.i.d. geometric random variables with parameter* $\frac{1}{1+v_i}$ *for any* l *and* $k = 1, 2$. *Therefore, they are unbiased i.i.d. estimators of* v_i.

Proof. Lemma 7 can be derived directly from Lemma 12 which is stated and shown next. ■

Lemma 12 *The moment generating function of the estimator conditioned on* S_1^l *and* S_2^l, $\hat{v}_{i,l}^{(k)}$, *is given by*

$$E_\pi\left[e^{\theta \hat{v}_{i,l}^{(k)}}\right] = \frac{1}{1 - v_i(e^\theta - 1)}$$

for all $i \in S_k^l$, $\theta \leq \log \frac{1+v_i}{v_i}$, *and* $k = 1, 2$.

Proof. Let N_k^l denote the number of customers who enter tier k in epoch $l \in \mathcal{L}_k$ before no purchase in S_k^l occurs. The probability of no purchase for S_k^l conditioned on entering tier k is

$$p_0(S_k^l) = \frac{1}{1 + \sum_{j \in S_k^l} v_j}.$$

Given any fixed value of N_k^l, $\hat{v}_{i,l}^{(k)}$ is a binomial random variable with N_k^l trials and the success probability (i.e., exiting without choosing any product from current tier) is given by

$$q_i^{(k)}(S_k^l) = P(\text{choose } i \text{ from } S_k^l | \text{choose at least one product})$$

$$= \frac{v_i}{1+\sum_{j\in S_k^l} v_j} \Bigg/ \left(1 - \frac{1}{1+\sum_{j\in S_k^l} v_j}\right) = \frac{v_i}{\sum_{j\in S_k^l} v_j}.$$

Since the moment generating function for a binomial random variable with parameters n, p is $(pe^\theta + 1 - p)^n$, we have

$$E_\pi\left[e^{\theta \hat{v}_{i,l}} | N_k^l\right] = E_{N_k^l}\left[(q_i^{(k)} e^\theta + 1 - q_i^{(k)})^{N_k^l}\right].$$

Therefore,

$$E_\pi\left[e^{\theta \hat{v}_{i,l}}\right] = E\left[E_{N_k^l}\left[(q_i^{(k)} e^\theta + 1 - q_i^{(k)})^{N_k^l}\right]\right]$$
$$= \frac{p_0(S_k^l)}{1 - (q_i^{(k)} e^\theta + 1 - q_i^{(k)})(1 - p_0(S_k^l))}$$
$$= \frac{1}{1 - v_i(e^\theta - 1)} \text{ for all } \theta < \log \frac{1+v_i}{v_i}.$$

■

Lemma 8 *Assume $0 \le v_i \le v_i^{UCB}$ for all $i = 1, \cdots, K$. Suppose $\mathbf{S}^*$ is an optimal tiered recommendation when the parameters of SMNL model are given by $\mathbf{v}$. Then $E[R(\mathbf{S}^*, \mathbf{v}^{UCB})] \ge E[R(\mathbf{S}^*, \mathbf{v})]$.*

Proof. For any product $i \in S_2^*$, we have $r_i \ge E[R(S_2^*, \mathbf{v})]$ where

$$E[R(S_2^*, \mathbf{v})] = \frac{\sum_{j\in S_2^*} r_j v_j}{1+\sum_{j\in S_2^*} v_j},$$

otherwise removing it from S_2^* will improve the expected profit. Let $\Delta v_i = v_i^{UCB} - v_i$, then we have

$$E[R(S_2^*, \mathbf{v}^{UCB})] = \frac{\sum_{j\in S_2^*} r_j v_j + \sum_{j\in S_2^*} r_j \Delta v_j}{1+\sum_{j\in S_2^*} v_j + \sum_{j\in S_2^*} \Delta v_j}.$$

Since $r_j \ge E[R(S_2^*, \mathbf{v})]$ for $j \in S_2^*$, we have

$$\sum_{j\in S_2^*} r_j \Delta v_j \ge \sum_{j\in S_2^*} \Delta v_j E[R(S_2^*, \mathbf{v})],$$

which implies that

$$\sum_{j\in S_2^*} r_j \Delta v_j \left(1 + \sum_{j\in S_2^*} v_j\right) \ge \left(\sum_{j\in S_2^*} r_j v_j\right)\left(\sum_{j\in S_2^*} \Delta v_j\right).$$

Adding $\left(1+\sum_{j\in S_2^*} v_j\right)\left(\sum_{j\in S_2^*} r_j v_j\right)$ to both sides,

$$\sum_{j\in S_2^*} r_j \Delta v_j \left(1 + \sum_{j\in S_2^*} v_j\right) + \left(1 + \sum_{j\in S_2^*} v_j\right)\left(\sum_{j\in S_2^*} r_j v_j\right)$$

$$\geq \left(\sum_{j\in S_2^*} r_j v_j\right)\left(\sum_{j\in S_2^*}\Delta v_j\right) + \left(1+\sum_{j\in S_2^*} v_j\right)\left(\sum_{j\in S_2^*} r_j v_j\right) = \left(\sum_{j\in S_2^*} r_j v_j\right)\left(1+\sum_{j\in S_2^*} v_j + \sum_{j\in S_2^*}\Delta v_j\right).$$

Therefore, we obtain the following

$$\left(\sum_{j\in S_2^*} r_j v_j + \sum_{j\in S_2^*} r_j \Delta v_j\right)\left(1+\sum_{j\in S_2^*} v_j\right) \geq \left(\sum_{j\in S_2^*} r_j v_j\right)\left(1+\sum_{j\in S_2^*} v_j + \sum_{j\in S_2^*}\Delta v_j\right).$$

It implies that

$$E[R(S_2^*,\mathbf{v}^{UCB})] = \frac{\sum_{j\in S_2^*} r_j v_j + \sum_{j\in S_2^*} r_j\Delta v_j}{1+\sum_{j\in S_2^*} v_j + \sum_{j\in S_2^*}\Delta v_j} \geq \frac{\sum_{j\in S_2^*} r_j v_j}{1+\sum_{j\in S_2^*} v_j} = E[R(S_2^*,\mathbf{v})].$$

Similarly we can prove that

$$\frac{\sum_{j\in S_1^*} r_j v_j + \sum_{j\in S_1^*} r_j\Delta v_j + E[R(S_2^*,\mathbf{v})]}{1+\sum_{j\in S_1^*} v_j + \sum_{j\in S_1^*}\Delta v_j} \geq \frac{\sum_{j\in S_1^*} r_j v_j + E[R(S_2^*,\mathbf{v})]}{1+\sum_{j\in S_1^*} v_j} = E[R(\mathbf{S}^*,\mathbf{v})]$$

for $\Delta v_j > 0$. Combining both results, we have

$$\begin{aligned} E[R(\mathbf{S}^*,\mathbf{v}^{UCB})] &= \frac{\sum_{j\in S_1^*} r_j v_j + \sum_{j\in S_1^*} r_j\Delta v_j + E[R(S_2^*,\mathbf{v}^{UCB})]}{1+\sum_{j\in S_1^*} v_j + \sum_{j\in S_1^*}\Delta v_j} \\ &\geq \frac{\sum_{j\in S_1^*} r_j v_j + \sum_{j\in S_1^*} r_j\Delta v_j + E[R(S_2^*,\mathbf{v})]}{1+\sum_{j\in S_1^*} v_j + \sum_{j\in S_1^*}\Delta v_j} \geq E[R(\mathbf{S}^*,\mathbf{v})]. \end{aligned}$$

■

The following lemma is a straightforward derivation from Corollary D.1 from Agrawal et al., 2017a.

Lemma 13 (Concentration bound of geometric random variable) *For n i.i.d. geometric random variables $X_1,\cdots,X_n$ with parameter $\frac{1}{1+v_i}$ with $v_i \leq 1$, we have*

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - v_i\right| > \sqrt{\frac{48\log(\lambda+1)}{n}} + \frac{48\log(\lambda+1)}{n}\right) \leq \frac{4}{\lambda^2}$$

for any $\lambda > 0$ and all $n \in \mathbb{N}^+$.

Lemma 9 (Minimum learning criterion) *For any ϵ and $\alpha > 0$, if the number of epochs $M \geq \frac{192\log(2/\alpha+1)}{(-1+\sqrt{1+4\epsilon})^2}$, then $\bar{v}_i$ is within the ϵ-confidence bound of v_i with probability at least $1-\alpha$. That is,*

$$P(|\bar{v}_{i,l} - v_i| > \epsilon) < 1-\alpha$$

if $T_i(l) > \frac{192\log(2/\alpha+1)}{(-1+\sqrt{1+4\epsilon})^2}$.

Proof. Set $\lambda = 2/\alpha$. Since $f(t) = t^2 + t - \epsilon$ is negative on $[\frac{-1-\sqrt{-1+4\epsilon}}{2}, \frac{-1+\sqrt{-1+4\epsilon}}{2}]$ and

$$0 < \sqrt{\frac{48\log(\lambda+1)}{M}} \leq \frac{-1+\sqrt{-1+4\epsilon}}{2},$$

we have

$$\sqrt{\frac{48\log(\lambda+1)}{M}} + \frac{48\log(\lambda+1)}{M} \leq \epsilon.$$

Therefore, applying Lemma 13, we have

$$P\left(|\hat{v}_i - v_i| > \epsilon\right) \leq P\left(|\hat{v}_i - v_i| > \sqrt{\frac{48\log(\lambda+1)}{M}} + \frac{48\log(\lambda+1)}{M}\right) \leq \frac{4}{\lambda^2} = \alpha.$$

Therefore,

$$P\left(|\hat{v}_i - v_i| < \epsilon\right) \geq 1 - \alpha.$$

■

Lemma 14 *The expected profits during one learning epoch, conditioned on* $\mathbf{S}'$, $\mathbf{S}''$, *and* $\mathbf{v}$, *for the two strategies are*

$$E\left[\sum_{t=1}^{N_1} R_t(\mathbf{S}')\right] = \sum_{j\in S_1} r_j v_j + r_m v_m + \frac{\sum_{j\in S_2} r_j v_j}{1+\sum_{j\in S_2} v_j}$$

and

$$E\left[\sum_{t=1}^{N_2} R_t(\mathbf{S}'')\right] = \left(1 + \sum_{j\in S_2} v_j + v_m\right)\sum_{j\in S_1} r_j v_j + \sum_{j\in S_2} r_j v_j + r_m v_m.$$

Proof. Conditioned on N_1, the probability of choosing product $j \in S_1'$ at the time before N_1 is

$$\frac{v_j}{1+\sum_{i\in S_1'} v_i} \Big/ \left(1 - \frac{1}{1+\sum_{i\in S_1'} v_i}\right),$$

and the expected profit to obtain at time N_1 is $\frac{\sum_{j\in S_2} r_j v_j}{1+\sum_{j\in S_2} v_j}$. Thus the expected profit obtained during the process of getting one sample is

$$\begin{aligned}
&E\left[\sum_{t=1}^{N_1} R_t(\mathbf{S}')\right] = E\left[E\left[\sum_{t=1}^{N_1} R_t(\mathbf{S}')\middle| N_1\right]\right] \\
&= E\left[\sum_{i=1}^{N_1-1} \frac{1}{1+\sum_{j\in S_1} v_j + v_m}\left(\sum_{j\in S_1} r_j v_j + r_m v_m\right)\Big/\left(1 - \frac{1}{1+\sum_{j\in S_1} v_j + v_m}\right)\right] + \frac{\sum_{j\in S_2} r_j v_j}{1+\sum_{j\in S_2} v_j} \\
&= E\left[\sum_{i=1}^{N_1-1} \frac{1}{\sum_{j\in S_1} v_j + v_m}\left(\sum_{j\in S_1} r_j v_j + r_m v_m\right)\right] + \frac{\sum_{j\in S_2} r_j v_j}{1+\sum_{j\in S_2} v_j}
\end{aligned}$$

$$= \sum_{j\in S_1} r_j v_j + r_m v_m + \frac{\sum_{j\in S_2} r_j v_j}{1+\sum_{j\in S_2} v_j},$$

where the last equality follows from the fact that N_1 is a geometric random variable with mean $1+\sum_{j\in S_1} v_j + v_m$ so $E[N_1-1] = \sum_{j\in S_1} v_j + v_m$. Alternatively, since N_1 is a stopping time, the above result can be computed through Wald's equation

$$E\left[\sum_{t=1}^{N_1} R_t(\mathbf{S}')\right] = E[N_1]E[R(\mathbf{S}')] = \sum_{j\in S_1} r_j v_j + r_m v_m + \frac{\sum_{j\in S_2} r_j v_j}{1+\sum_{j\in S_2} v_j}.$$

If we add the new product to S_2. Then the expected profit gained during the process of one sample obtained is

$$\begin{aligned}
E\left[\sum_{t=1}^{N_2} R_t(\mathbf{S}'')\right] &= E\left[E\left[\sum_{t=1}^{N_2-1} R_t(\mathbf{S}'')\middle| N_2\right]\right] \\
&= E\left[\sum_{t=1}^{N_2-1} \frac{1}{1+\sum_{j\in S_1} v_j}\left(\sum_{j\in S_1} r_j v_j + \frac{\sum_{j\in S_2} r_j v_j + r_m v_m}{1+\sum_{j\in S_2} v_j + v_m}\right)\Big/ \right. \\
&\quad \left.\left(1 - \frac{1}{(1+\sum_{j\in S_1} v_j)(1+\sum_{j\in S_2} v_j + v_m)}\right)\right] \\
&= \left(1+\sum_{j\in S_2} v_j + v_m\right)\left(\sum_{j\in S_1} r_j v_j + \frac{\sum_{j\in S_2} r_j v_j + r_m v_m}{1+\sum_{j\in S_2} v_j + v_m}\right) \\
&= \left(1+\sum_{j\in S_2} v_j + v_m\right)\sum_{j\in S_1} r_j v_j + \sum_{j\in S_2} r_j v_j + r_m v_m,
\end{aligned}$$

where the second last equality follows from the fact that N_2 is a geometric random random variable with mean $\left(1+\sum_{j\in S_1} v_j\right)\left(1+\sum_{j\in S_2} v_j + v_m\right)$. ■

Theorem 10 *The optimal solution to $G^{(1)}$ and $G^{(2)}$ is the same as $\mathbf{S}^*$. That is,*

$$\mathbf{Q}^* = \mathbf{Q}'^* = \mathbf{S}^*.$$

In addition, we have

$$G^{(1)}(\mathbf{S}^*, \mathbf{v}) \geq G^{(2)}(\mathbf{S}^*, \mathbf{v}) = v_m(E[R(S_2^*)] - r_m).$$

Proof. Let $\mathbf{Q}^*$ and $\mathbf{Q}'^*$ denote the optimal tiered recommendation solution that minimizes the regret $G^{(1)}$ and $G^{(2)}$, respectively. Note that

$$G^{(1)}(\mathbf{Q}, \mathbf{v}) = E[R(\mathbf{S}^*)]\left(1+\sum_{j\in Q_1} v_j + v_m\right) - \left(\sum_{j\in Q_1} r_j v_j + r_m v_m + \frac{\sum_{j\in Q_2} r_j v_j}{1+\sum_{j\in Q_2} v_j}\right).$$

Let $\mathbf{Q}^*$ denote the optimal solution. The optimal recommendation $\mathbf{Q}^* = (Q_1^*, Q_2^*)$ satisfies

$$Q_1^* = \underset{Q_1}{\text{argmin}} \sum_{j\in Q_1} v_j(E[R(\mathbf{S}^*)] - r_j),$$

and

$$Q_2^* = \underset{Q_2}{\text{argmin}} -\frac{\sum_{j\in Q_2} r_j v_j}{1+\sum_{j\in Q_2} v_j}.$$

From the above equation, $j \in Q_1^*$ if and only if $r_j > E[R(\mathbf{S}^*)]$. As is shown in Proposition 4, it implies that $Q_1^* = S_1^*$. The second equality implies that $Q_2^* = S_2^*$ so $\mathbf{Q}^* = \mathbf{S}^*$.

Note that

$$\begin{aligned} G^{(2)}(\mathbf{Q}', \mathbf{v}) = E[R(\mathbf{S}^*)] &\left(1 + \sum_{j\in Q_1'} v_j\right)\left(1 + \sum_{j\in Q_2'} v_j + v_m\right) \\ &- \left(1+\sum_{j\in Q_2'} v_j + v_m\right)\sum_{j\in Q_1'} r_j v_j - \sum_{j\in Q_2'} r_j v_j - r_m v_m. \end{aligned}$$

Minimizing $G^{(2)}(\mathbf{Q}', \mathbf{v})$ is equivalent to solving the optimization problem

$$\max_{\mathbf{Q}'} \left(1+\sum_{j\in Q_2'} v_j\right)\left(1+\sum_{j\in Q_1'} v_j\right)(E[R(\mathbf{Q}')]-E[R(\mathbf{S}^*)])+v_m\left(1+\sum_{j\in Q_1'} v_j\right)(E[R(Q_1')]-E[R(\mathbf{S}^*)]). \tag{1}$$

For any fixed set Q_2', to find the optimal Q_1', note that

$$\begin{aligned} Eq\ (1) = &\left(1+\sum_{j\in Q_2'} v_j\right)\left(\sum_{j\in Q_1'} r_j v_j + E[R(Q_2')]\right) + v_m\left(\sum_{j\in Q_1'} r_j v_j + E[R(Q_2')]\right) \\ &- \left(1+\sum_{j\in Q_2'} v_j + v_m\right)\left(1+\sum_{j\in Q_1'} v_j\right)E[R(\mathbf{S}^*)] \\ = &\left(1+\sum_{j\in Q_2'} v_j + v_m\right)\sum_{j\in Q_1'} r_j v_j - E[R(\mathbf{S}^*)]\left(1+\sum_{j\in Q_2'} v_j + v_m\right)\left(\sum_{j\in Q_1'} v_j\right) \\ &+ (E[R(Q_2')] - E[R(\mathbf{S}^*)])\left(1+\sum_{j\in Q_2'} v_j + v_m\right). \end{aligned}$$

Therefore

$$Q_1'^* = \max_{Q_1'} \quad \left(1+\sum_{j\in Q_2'} v_j + v_m\right)\sum_{j\in Q_1'} r_j v_j - E[R(\mathbf{S}^*)]\left(1+\sum_{j\in Q_2'} v_j + v_m\right)\left(\sum_{j\in Q_1'} v_j\right)$$

$$= \max_{Q_1'} \sum_{j\in Q_1'} r_j v_j - E[R(\mathbf{S}^*)] \left(\sum_{j\in Q_1'} v_j\right).$$

It implies that $j \in Q_1'^*$ if and only if $r_j \geq E[R(\mathbf{S}^*)]$. For any $i \in S_1^*$, we have $r_i \geq E[R(\mathbf{S}^*)]$, which implies that $S_1^* \subset Q_1'^*$. Also, for any product that $r_i \geq E[R(\mathbf{S}^*)]$ and $i \in X_1$, it is included in $Q_1'^*$. It implies that the optimal $Q_1'^*$ is S_1^*. For the fixed set S_1^*, we find the optimal Q_2'. Note that the second term in Eq 1 is not dependent on Q_2', and the first term with $Q_1' = S_1^*$ equals

$$\begin{aligned}
&\left(1+\sum_{j\in Q_2'} v_j\right)\left(1+\sum_{j\in S_1^*} v_j\right)(E[R((S_1^*,Q_2'))] - E[R(\mathbf{S}^*)])\\
&= \left(1+\sum_{j\in Q_2'} v_j\right)\left(\sum_{j\in S_1^*} r_j v_j\right) + \sum_{j\in Q_2'} r_j v_j - \left(1+\sum_{j\in Q_2'} v_j\right)\left(1+\sum_{j\in S_1^*} v_j\right) E[R(\mathbf{S}^*)]\\
&= \left(\sum_{j\in S_1^*} r_j v_j - \left(1+\sum_{j\in S_1^*} v_j\right) E[R(\mathbf{S}^*)]\right)\left(\sum_{j\in Q_2'} v_j\right) + \sum_{j\in Q_2'} r_j v_j\\
&\quad + \sum_{j\in S_1^*} r_j v_j - \left(1+\sum_{j\in S_1^*} v_j\right) E[R(\mathbf{S}^*)].
\end{aligned}$$

For the optimal $Q_2'^*$,

$$\begin{aligned}
Q_2'^* &= \operatorname*{argmax}_{Q_2'} \left(\sum_{j\in S_1^*} r_j v_j - \left(1+\sum_{j\in S_1^*} v_j\right) E[R(\mathbf{S}^*)]\right)\left(\sum_{j\in Q_2'} v_j\right) + \sum_{j\in Q_2'} r_j v_j\\
&= \operatorname*{argmax}_{Q_2'} \left(1+\sum_{j\in S_1^*} v_j\right)(E[R(S_1^*)] - E[R(\mathbf{S}^*)])\left(\sum_{j\in Q_2'} v_j\right) + \sum_{j\in Q_2'} r_j v_j.
\end{aligned}$$

Therefore $j \in Q_2'$ if and only if

$$r_j > \left(1+\sum_{j\in S_1^*} v_j\right)(E[R(\mathbf{S}^*)] - E[R(S_1^*)]) = E[R(S_2^*)].$$

It implies that $Q_2'^* = S_2^*$ so $\mathbf{Q}'^* = \mathbf{S}^*$.

For $G^{(1)}(\mathbf{S}^*, \mathbf{v})$, we have

$$\begin{aligned}
G^{(1)}(\mathbf{S}^*, \mathbf{v}) &= E[R(\mathbf{S}^*)]\left(1+\sum_{j\in S_1^*} v_j + v_m\right) - \left(\sum_{j\in S_1^*} r_j v_j + r_m v_m + \frac{\sum_{j\in S_2^*} r_j v_j}{1+\sum_{j\in S_2^*} v_j}\right)\\
&= E[R(\mathbf{S}^*)]\left(1+\sum_{j\in S_1^*} v_j\right) - \left(\sum_{j\in S_1^*} r_j v_j + \frac{\sum_{j\in S_2^*} r_j v_j}{1+\sum_{j\in S_2^*} v_j}\right) + E[R(\mathbf{S}^*)] v_m - v_m r_m
\end{aligned}$$

$$= v_m(E[R(\mathbf{S}^*)] - r_m).$$

Similarly, we have

$$\begin{aligned}
G^{(2)}(\mathbf{S}^*, \mathbf{v}) &= E[R(\mathbf{S}^*)]\left(1 + \sum_{j\in S_1^*} v_j\right)\left(1 + \sum_{j\in S_2^*} v_j + v_m\right) \\
&\quad - \left(1 + \sum_{j\in S_2^*} v_j + v_m\right)\sum_{j\in S_1^*} r_j v_j - \sum_{j\in S_2^*} r_j v_j - r_m v_m \\
&= \left(1 + \sum_{j\in S_2^*} v_j\right)\left(E[R(\mathbf{S}^*)]\left(1 + \sum_{j\in S_1^*} v_j\right) - \sum_{j\in S_1^*} r_j v_j - \frac{\sum_{j\in S_2^*} r_j v_j}{1 + \sum_{j\in S_2^*} v_j}\right) \\
&\quad + v_m\left(1 + \sum_{j\in S_1^*} v_j\right)\left(E[R(\mathbf{S}^*)] - \frac{\sum_{j\in S_1^*} r_j v_j}{1 + \sum_{j\in S_1^*} v_j}\right) - r_m v_m \\
&= v_m(E[R(S_2^*)] - r_m).
\end{aligned}$$

Since $E[R(\mathbf{S}^*)] \geq E[R(S_2^*)]$ as shown in Proposition 4, we have $G^{(1)}(\mathbf{S}^*, \mathbf{v}) \geq G^{(2)}(\mathbf{S}^*, \mathbf{v})$. ∎

Lemma 15 provides the concentration bound of $v_{i,l}^{UCB}$, which is a straightforward conclusion from Lemma 4.1 from Agrawal et al., 2017a.

Lemma 15 (Concentration bound) *For every $l \in \mathcal{L}$, we have*

a.) $v_{i,l}^{UCB} \geq v_i$ with probability at least $1 - \frac{6}{K^2 l}$ for all $i = 1, \cdots, K$.

b.) There exists constants C_1 and C_2 such that $v_{i,l}^{UCB} - \bar{v}_{i,l} \leq C_1\sqrt{\frac{v_i \log(Kl+1)}{T_i(l)}} + C_2\frac{\log(Kl+1)}{T_i(l)}$. with probability at least $1 - \frac{7}{K^2 l}$.

Theorem 11 (Performance bound for Algorithm 1) *The regret during time $[0, T]$ is bounded above by*

$$\begin{aligned}
Reg_\pi(T; \mathbf{v}) \leq CK\log^2(KT) + C\sqrt{TK\log(KT)} \\
+ M\sum_{i\in X} v_i(r_{max} - r_i),
\end{aligned}$$

for some constant C where r_{max} is the highest profit among X, and K is the total number of products.

Proof. Since products are being dynamically launched, the optimal recommendation is changing. Let $\mathbf{S}_l^*$ ($\tilde{\mathbf{S}}^l$) be the optimal recommendation among sets (X_1^l, X_2^l) with value $\mathbf{v}$ ($\mathbf{v}_l^{UCB}$). The new product is defined as the product whose learning epoch is less than M. Define H_l as the set of new products that are not in $\tilde{S}_2^l$ but are added to the second-tier recommendation at epoch l for $l \in \mathcal{L}_2$.

Also define the function $\kappa(l)$ as a set of epochs on tier 1 which corresponds to epoch $l \in \mathcal{L}_2$. (In the example of Fig 1, $\kappa(0) = \{0, 1\}$, $\kappa(3) = \{3, 4\}$.) The summation of the regret until time T is

$$Reg_\pi(T; \mathbf{v}) = E_\pi \left\{ \sum_{l \in \mathcal{L}_2} \sum_{j \in \kappa(l)} \sum_{t \in \varepsilon_j^1} \left(R_t(\mathbf{S}_j^*, \mathbf{v}) - R_t\left((\tilde{S}_1^j, \tilde{S}_2^l \cup H_l), \mathbf{v} \right) \right) \right\},$$

where R_t denotes the profit obtained at time t. Define $\mathcal{F}_l$ as the filtration associated with the policy up to epoch l. Since

$$\begin{aligned} &E_\pi \left\{ \sum_{l \in \mathcal{L}_2} \sum_{j \in \kappa(l)} \sum_{t \in \varepsilon_j^1} R_t\left((\tilde{S}_1^j, (\tilde{S}_2^l \cup H_l)), \mathbf{v} \right) \right\} \\ &= E_\pi \left\{ \sum_{l \in \mathcal{L}_2} \sum_{j \in \kappa(l)} \sum_{t \in \varepsilon_j^1} R_t\left(\tilde{S}_1^j, \mathbf{v} \right) + \sum_{l \in \mathcal{L}_2} \sum_{t \in \varepsilon_j^2} R_t(\tilde{S}_2^l \cup H_l, \mathbf{v}) \right\}, \end{aligned}$$

we have

$$Reg_\pi(T; \mathbf{v}) = E_\pi \left\{ \sum_{l \in \mathcal{L}_2} \left(\left(\sum_{j \in \kappa(l)} \sum_{t \in \varepsilon_j^1} R_t(\mathbf{S}_j^*, \mathbf{v}) - R_t(\tilde{S}_1^j, \mathbf{v}) \right) - \sum_{t \in \varepsilon_2^l} R_t(\tilde{S}_2^l \cup H_l, \mathbf{v}) \right) \right\}.$$

For all $l = 1, \cdots, |\mathcal{L}|$, define the event A_l as

$$A_l = \bigcap_{i=1}^{K} \left\{ v_{i,l}^{UCB} - C_1 \sqrt{\frac{v_i \log(K(l - l_{i,0}) + 1)}{T_i(l)}} + C_2 \frac{\log(K(l - l_{i,0}) + 1)}{T_i(l)} < v_i < v_{i,l}^{UCB} \right\}.$$

For products not launched before epoch l, we assume they directly fall into this confidence bound in epoch l. Conditional on $A_j A_l$ for $j \in \mathcal{L}_1$ and $l \in \mathcal{L}_2$, according to Lemma 8, we have

$$E[R(\tilde{\mathbf{S}}^l, \mathbf{v})] \le E[R(\mathbf{S}^*, \mathbf{v})] \le E[R(\mathbf{S}^*, \mathbf{v}_l^{UCB})] \le E[R(\tilde{\mathbf{S}}^l, \mathbf{v}_l^{UCB})],$$

so $E[R(\mathbf{S}^*, \mathbf{v})] - E[R(\tilde{\mathbf{S}}^l, \mathbf{v})]$ can be bounded from above by $E_\pi[R(\tilde{\mathbf{S}}_j, \mathbf{v}_j^{UCB} \oplus \mathbf{v}_l^{UCB})] - E_\pi[R(\mathbf{S}_j^*, \mathbf{v})]$ where $\mathbf{v}_j^{UCB} \oplus \mathbf{v}_l^{UCB}$ denotes using $\mathbf{v}_j^{UCB}$ and $\mathbf{v}_l^{UCB}$ to calculate the first- and second-tier recommendation, respectively. Then we have for a given $l \in \mathcal{L}_2$,

$$\begin{aligned} &E_\pi \left\{ \left(\left(\sum_{j \in \kappa(l)} \sum_{t \in \varepsilon_j^1} R_t(\mathbf{S}_j^*, \mathbf{v}) - R_t(\tilde{S}_1^j, \mathbf{v}) \right) - \sum_{t \in \varepsilon_2^l} R_t(\tilde{S}_2^l \cup H_l, \mathbf{v}) \right) 1(A_l) \right\} \\ &\le E_\pi \left\{ \sum_{j \in \kappa(l)} \left(\sum_{t \in \varepsilon_j^1} R_t(\mathbf{S}_j^*, \mathbf{v}) - R_t(\tilde{S}_1^j, \mathbf{v}) \right) 1(A_j A_l) - \sum_{t \in \varepsilon_l^2} R_t(\tilde{S}_2^l \cup H_l, \mathbf{v}) 1(A_l) + \sum_{j \in \kappa(l)} 1(A_j^c) \sum_{t \in \varepsilon_j^1} r_{max} \right\} \\ &\le E_\pi \left\{ \sum_{j \in \kappa(l)} \left(\sum_{t \in \varepsilon_j^1} R_t(\tilde{\mathbf{S}}^j, \mathbf{v}_j^{UCB} \oplus \mathbf{v}_l^{UCB}) - R_t(\tilde{S}_1^j, \mathbf{v}) \right) 1(A_j A_l) \right. \end{aligned}$$

$$-\sum_{t\in\varepsilon_l^2} R_t(\tilde{S}_2^l \cup H_l, \mathbf{v})1(A_l) + \sum_{j\in\kappa(l)} 1(A_j^c) E_\pi\left[\left[\sum_{t\in\varepsilon_j^1} r_{max}|\mathcal{F}_{j-1}\right]\right]\Bigg\}$$

$$\leq E_\pi\left\{\sum_{j\in\kappa(l)} 1(A_jA_l) E\left[\sum_{t\in\varepsilon_j^1} R_t(\tilde{\mathbf{S}}^j, \mathbf{v}_j^{UCB}\oplus\mathbf{v}_l^{UCB}) - R_t(\tilde{S}_1^j, \mathbf{v})\middle|\mathcal{F}_{j-1}\right]\right.$$

$$\left. -\sum_{t\in\varepsilon_l^2} R_t(\tilde{S}_2^l \cup H_l, \mathbf{v})1(A_l) + r_{max}(1+K)\sum_{j\in\kappa(l)} 1(A_j^c)\right\},$$

where r_{max} is the maximum profit among all products in X. Since on the event A_jA_l, we have

$$E\left[\sum_{t\in\varepsilon_j^1} R_t(\tilde{\mathbf{S}}^j, \mathbf{v}_j^{UCB}\oplus\mathbf{v}_l^{UCB}) - R_t(\tilde{S}_1^j, \mathbf{v})\middle|\mathcal{F}_{j-1}\right]$$

$$= \left(1+\sum_{i\in\tilde{S}_1^j} v_i\right)\frac{\sum_{i\in\tilde{S}_1^j} r_i v_{i,j}^{UCB} + \frac{\sum_{i\in\tilde{S}_2} r_i v_{i,l}^{UCB}}{1+\sum_{i\in\tilde{S}_2} v_{i,l}^{UCB}}}{1+\sum_{i\in\tilde{S}_1^j} v_{i,j}^{UCB}} - \sum_{i\in\tilde{S}_1^j} r_i v_i$$

$$\leq \sum_{i\in\tilde{S}_1^j} r_i(v_{i,j}^{UCB} - v_i) + E[R(\tilde{S}_2^l, \mathbf{v}_l^{UCB})|\mathcal{F}_{l-1}],$$

where the first equality uses Wald's equation and the fact that $|\epsilon_j^1|$ follows the geometric distribution with mean $(1+\sum_{i\in\tilde{S}_1^j} v_i)$. Therefore, we have

$$\begin{aligned}
Reg_\pi(T;\mathbf{v}) \leq E_\pi\Bigg\{&\sum_{l\in\mathcal{L}_2}\sum_{j\in\kappa(l)}\sum_{i\in\tilde{S}_1^j} r_i(v_{i,j}^{UCB} - v_i)1(A_j)\\
&+\sum_{l\in\mathcal{L}_2}\sum_{t\in\varepsilon_l^2}\Big(E[R_t(\tilde{S}_2^l, \mathbf{v}_l^{UCB})|\mathcal{F}_{l-1}] - R_t(\tilde{S}_2^l\cup H_l, \mathbf{v})\Big)1(A_l)\\
&+r_{max}(1+K)\sum_{l\in\mathcal{L}_2}\sum_{j\in\kappa(l)} 1(A_j^c) + r_{max}\sum_{l\in\mathcal{L}_2}\sum_{j\in\kappa(l)}\sum_{t\in\varepsilon_j^1} 1(A_l^c)\Bigg\}\\
\leq E_\pi\Bigg\{&\sum_{l\in\mathcal{L}_2}\sum_{j\in\kappa(l)}\sum_{i\in\tilde{S}_1^j} r_i(v_{i,j}^{UCB} - v_i)1(A_j)\\
&+\sum_{l\in\mathcal{L}_2}\sum_{t\in\varepsilon_l^2}\Big(E[R_t(\tilde{S}_2^l, \mathbf{v}_l^{UCB})|\mathcal{F}_{l-1}] - R_t(\tilde{S}_2^l\cup H_l, \mathbf{v})\Big)1(A_l)\\
&+r_{max}(1+K)\sum_{j\in\mathcal{L}_1} 1(A_j^c) + r_{max}(1+K)^2\sum_{l\in\mathcal{L}_2} 1(A_l^c)\Bigg\}.
\end{aligned}$$

Since

$$E_\pi \left\{ \sum_{l \in \mathcal{L}_2} \sum_{t \in \varepsilon_l^2} \Big(E[R_t(\tilde{S}_2^l, \mathbf{v}_l^{UCB}) | \mathcal{F}_{l-1}] - R_t(\tilde{S}_2^l \cup H_l, \mathbf{v}) \Big) 1(A_l) \right\}$$

$$= E_\pi \left\{ \sum_{l \in \mathcal{L}_2} 1(A_l) E\left[\sum_{t \in \varepsilon_l^2} \Big(R_t(\tilde{S}_2^l, \mathbf{v}_l^{UCB}) - R_t(\tilde{S}_2^l \cup H_l, \mathbf{v}) \Big) \middle| \mathcal{F}_{l-1} \right] \right\}$$

$$\leq E_\pi \left\{ \sum_{l \in \mathcal{L}_2} \left(1 + \sum_{i \in \tilde{S}_2^l} v_i + \sum_{i \in H_l} v_i \right) \Big(E[R(\tilde{S}_2^l, \mathbf{v}_l^{UCB}) - R(\tilde{S}_2^l \cup H_l, \mathbf{v}) | \mathcal{F}_{l-1}] \Big) 1(A_l) \right\}$$

$$= E_\pi \left\{ \sum_{l \in \mathcal{L}_2} \left(\sum_{i \in \tilde{S}_2^l} r_i (v_{i,l}^{UCB} - v_i) + \sum_{i \in H_l} v_i (E[R(\tilde{S}_2^l, \mathbf{v}_l^{UCB})] - r_i) \right) 1(A_l) \right\}$$

$$= E_\pi \left[\sum_{l \in \mathcal{L}_2} \sum_{i \in \tilde{S}_2^l} r_i (v_{i,l}^{UCB} - v_i) 1(A_l) \right] + E_\pi \left[\sum_{l \in \mathcal{L}_2} \sum_{i \in H_l} v_i (E[R(\tilde{S}_2^l, \mathbf{v}_l^{UCB})] - r_i) 1(A_l) \right],$$

and

$$E_\pi \left[\sum_{l \in \mathcal{L}_2} \sum_{i \in H_l} v_i (E[R(\tilde{S}_2^l, \mathbf{v}_l^{UCB})] - r_i) 1(A_l) \right] \leq M \sum_{i \in X} v_i (r_{max} - r_i).$$

The above inequality holds since any product will be included in H for at most M times. Thus, we have

$$Reg_\pi(T; \mathbf{v}) \leq E_\pi \left[\Delta R + r_{max}(1+K) \sum_{l \in \mathcal{L}_1} 1(A_l^c) + r_{max}(1+K)^2 \sum_{l \in \mathcal{L}_2} 1(A_l^c) \right] + M \sum_{i \in X} v_i (r_{max} - r_i),$$

where

$$\Delta R = \sum_{l \in \mathcal{L}_1} \sum_{i \in \tilde{S}_1^l} r_i (v_{i,l}^{UCB} - v_i) 1(A_l) + \sum_{l \in \mathcal{L}_2} \sum_{i \in \tilde{S}_2^l} r_i (v_{i,l}^{UCB} - v_i) 1(A_l).$$

According to Lemma 15, we have

$$P(A_l^c) \leq \sum_{i \in X} \frac{13}{K^2 (l - l_{i,0})} 1(l > l_{i,0}),$$

which implies that

$$E_\pi \left[(1+K) \sum_{l \in \mathcal{L}_1} 1(A_l^c) + (1+K)^2 \sum_{l \in \mathcal{L}_2} 1(A_l^c) \right]$$

$$\leq E_\pi \left[(1+K) \sum_{l \in \mathcal{L}_1} \sum_{i \in X} \frac{13}{K^2 (l - l_{i,0})} 1(l > l_{i,0}) + (1+K)^2 \sum_{l \in \mathcal{L}_2} \sum_{i \in X} \frac{13}{K^2 (l - l_{i,0})} 1(l > l_{i,0}) \right]$$

$$\leq \left(\frac{1+K}{K^2}+\frac{(1+K)^2}{K^2}\right)\sum_{t=1}^{2T} K\frac{13}{t} \leq C_3 K\log(T)$$

for some constant C_3. Since

$$\begin{aligned}\Delta R &\leq \sum_{l\in\mathcal{L}_2}\sum_{i\in\tilde{S}_2^l}\left(C_1\sqrt{\frac{v_i\log(K(l-l_{i,0})+1)}{T_i(l)}}+C_2\frac{\log(K(l-l_{i,0})+1)}{T_i(l)}\right)\\ &\quad+\sum_{l\in\mathcal{L}_1}\sum_{i\in\tilde{S}_1^l}\left(C_1\sqrt{\frac{v_i\log(K(l-l_{i,0})+1)}{T_i(l)}}+C_2\frac{\log(K(l-l_{i,0})+1)}{T_i(l)}\right)\\ &\leq 2\sum_{i=1}^{K}\sum_{t=1}^{T_i}\left(C_1\sqrt{\frac{v_i\log(KT+1)}{t}}+C_2\frac{\log(KT+1)}{t}\right),\end{aligned}$$

where T_i is total number of epochs for product i at time T, we have

$$\begin{aligned}Reg_\pi(T;\mathbf{v}) &\leq 2E_\pi\left[\sum_{i=1}^{K}\sum_{t=1}^{T_i}\left(C_1\sqrt{\frac{v_i\log(KT+1)}{t}}+C_2\frac{\log(KT+1)}{t}\right)\right]\\ &\quad+r_{max}E_\pi\left[(1+K)\sum_{l\in\mathcal{L}_1}1(A_l^c)+(1+K)^2\sum_{l\in\mathcal{L}_2}1(A_l^c)\right]+M\sum_{i\in X}v_i(r_{max}-r_i)\\ &\leq C_4K\log^2(KT)+C_4\sum_{i=1}^{K}E_\pi\left[\sqrt{v_iT_i\log(KT)}\right]+C_4K\log T+M\sum_{i\in X}v_i(r_{max}-r_i)\\ &\leq C_4K\log^2(KT)+C_4\sum_{i=1}^{K}\sqrt{v_iE_\pi[T_i]\log(KT)}+C_4K\log T+M\sum_{i\in X}v_i(r_{max}-r_i).\end{aligned}$$

Since

$$E_\pi\left[\sum_{l\in\mathcal{L}_1}|\varepsilon_l^1|+\sum_{l\in\mathcal{L}_2}|\varepsilon_l^2|\right]\leq 4T,$$

and

$$\begin{aligned}E_\pi\left[\sum_{k=1}^{2}\sum_{l\in\mathcal{L}_k}|\varepsilon_l^k|\right] &= E_\pi\left[\sum_{k=1}^{2}\sum_{l\in\mathcal{L}_k}E\left[|\varepsilon_l^k|\,\middle|\,\mathcal{F}\right]\right]\\ &\geq E_\pi\left[\sum_{k=1}^{2}\sum_{l\in\mathcal{L}_k}\left(1+\sum_{i\in S_k^l}v_i\right)\right]\geq\sum_{i=1}^{K}E_\pi[T_i]v_i,\end{aligned}$$

where $\mathcal{F}$ is the filtration corresponding to the recommendation offered in each epoch. To obtain the worst-case bound, we have

$$\max \quad \sum_{i=1}^{K}\sqrt{v_iE[T_i]\log(KT)}$$

$$s.t. \quad \sum_{i=1}^{K} E[T_i] v_i \le 4T.$$

The maximum possible objective value is $2\sqrt{TK\log(KT)}$. Therefore we conclude that

$$Reg_\pi(T;\mathbf{v}) \le CK\log^2(KT) + C\sqrt{TK\log(KT)} + M\sum_{i\in X} v_i(r_{max} - r_i)$$

for some constant C. ■

Remark: The regret bound can be extended to multiple tiers. Similar to the proof for Theorem 11, define the function $\kappa_r(l)$ as a set of epochs on tier r which corresponds to epoch $l \in \mathcal{L}_{r+1}$. Then the summation of the regret until time T is

$$\begin{aligned} &Reg_\pi(T;\mathbf{v}) \\ &= E_\pi\left\{\sum_{l_W\in\mathcal{L}_W}\sum_{l_{W-1}\in\kappa_{W-1}(l_W)}\cdots\sum_{l_1\in\kappa_1(l_2)}\sum_{t\in\varepsilon_{l_1}^1}\left(R_t(\mathbf{S}_j^*,\mathbf{v}) - R_t\left((\tilde{S}_1^{l_1},\tilde{S}_2^{l_2},\cdots,\tilde{S}_W^{l_W}\cup H_{l_W}),\mathbf{v}\right)\right)\right\}, \end{aligned}$$

and the remaining analysis follows similarly.