Appendix A: More on the Perfect Information Setting

A.1. Proof of Proposition 1

We prove part (i) by induction. Fix $j,k \in \{0,1\}$, and consider $t=T$ as the base case. Because $\rho_j(p)\min\{g_{\max}, \beta\hat{g}+\lambda_k(a)\}$ is nondecreasing in $\hat{g}$, $V_T^{(j,k)}\big((\hat{g},\hat{y})\big)$ is nondecreasing in $\hat{g}$. Now, assume that $\hat{V}_{t+1}^{(j,k)}((\hat{g},\hat{y}))$ is nondecreasing in $\hat{g}$ for any $\hat{y}$. By the induction assumption and the fact that $\rho_j(p)\min(g_{\max}, \beta\hat{g}+\lambda_k(a))$ is nondecreasing, it follows that $\hat{V}_t^{(j,k)}((\hat{g},\hat{y}))$ is nondecreasing in $\hat{g}$.

To prove part (ii), we fix $\hat{g}_{t-1} \in [g_{\min}, g_{\max})$ and rewrite the dynamic program (10) as follows:

$$\hat{V}_t^{(j,k)}((\hat{g}_{t-1},\hat{y}_{t-1})) = \max_{\substack{a\in[a_{\min},a_{\max}]\\ p\in[p_{\min},p_{\max}]\\ Q\in\{\hat{y}_{t-1},\ldots,\bar{N}\}}} \mathbb{E}^{(j,k)}\left[p\min\{D_t,Q\} - hQ + \hat{V}_{t+1}^{(j,k)}\big((\hat{g}_t, \max\{0, Q-D_t\})\big)\right].$$

The preceding equation follows from a change of variable: q is replaced by $Q = q + \hat{y}_{t-1}$ (and accordingly introducing a lower bound $\hat{y}_{t-1}$ on Q). As $\hat{y}_{t-1}$ decreases, we obtain a relaxation at the right-hand side because the objective function remains the same but the lower bound of the constraint $Q \geq \hat{y}_t$ decreases. Therefore, the optimal value at the left-hand side cannot decrease, i.e., $\hat{V}_t^{(j,k)}\big((\hat{g}_{t-1},\hat{y}_{t-1})\big)$ is nonincreasing in $\hat{y}_{t-1}$.

Part (iii) holds because the ordering cost is normalized to zero and $\hat{y}_t + \hat{\pi}_t^q(\hat{\omega}_t) \geq 1$. The first observation implies that order-up-to levels, rather than order quantities, are payoff relevant. The second observation means that the optimal perfect information policy orders up to at least one inventory unit in any period t. Q.E.D.

A.2. Reducing Search Space for Dynamic Programming Solution

To solve the dynamic program (10), we jointly optimize over a, p and q. In this optimization problem, the search space for the order quantity q is $\{0,\ldots,\bar{N}\}$. It is possible to shrink this search space using the monotonicity of $\hat{V}_t^{(j,k)}\big((\hat{g},\hat{y})\big)$ at the carryover inventory $\hat{y}$ (Proposition 1(ii)). To do so, we focus on the problem of choosing the optimal order quantity, which is given by

$$\max_{q\in\{0,\ldots,\bar{N}\}} \mathbb{E}^{(j,k)}\left[p\min\{D_t,\hat{y}_{t-1}+q\} - a - h(\hat{y}_{t-1}+q) + \hat{V}_{t+1}^{(j,k)}\big((\hat{g}_t,\max\{0,\hat{y}_{t-1}+q-D_t\})\big)\right]. \quad \text{(A.1)}$$

In this problem, we solve for the optimal order quantity given fixed values of p and a. As a result, the probability distribution of the demand D_t and the induced goodwill stock $\hat{g}_t$ are also fixed. This makes the preceding optimization problem akin to a classical newsvendor problem, albeit with an additional term, $\hat{V}_{t+1}^{(j,k)}\big((\hat{g}_t,\max\{0,\hat{y}_{t-1}+q-D_t\})\big)$. Our next result leverages this structure—together with Proposition 1(ii)—to reduce the search space for q.

Lemma 2. *For problem* (A.1)*, it is without loss of optimality to restrict attention to order quantities* $q \in \{0,\ldots,\max\{0, Q_m^*(\hat{g}_t,p)-\hat{y}_{t-1}\}\}$*, where* $Q_m^*(\hat{g}_t,p)$ *denotes the unique optimal solution to* $\max_{Q\in\{0,\ldots,\bar{N}\}} \mathbb{E}^{(j,k)}[p\min\{D_t,Q\}-hQ]$*, and* D_t *follows a binomial distribution with* $\bar{N}$ *trials and a success probability of* $\rho_j(p)\hat{g}_t$*.*

Proof of Lemma 2. Note that $\hat{V}_{t+1}^{(j,k)}\big((\hat{g}_t, \max\{0, q+\hat{y}_{t-1}-D_t\})\big)$ is nonincreasing in q due to Proposition 1(ii) and the fact that $\max\{0, q+\hat{y}_{t-1}-D_t\}$ is nondecreasing function of q. Second, it follows that $\mathbb{E}^{(j,k)}\left[p\min\{D_t, \hat{y}_{t-1}+q\} - a - h(q+\hat{y}_{t-1})\right]$ is decreasing in q if $q+\hat{y}_{t-1} > Q_m^*(\hat{g}_t, p_t)$ because it consists of a weighted sum of concave functions and a linear function. Therefore, the objective function in (A.1) is decreasing in q for $q > \max\{0, Q_m^*(\hat{g}_t, p_t) - \hat{y}_{t-1}\}$. Finally, we note that the newsvendor solution $Q_m^*(\hat{g}_t, p)$ is unique (see, for instance, Levi et al. 2015) and given by $Q_m^*(\hat{g}_t, p) = \inf\{Q : \mathbb{P}^{(j,k)}\{D_t \le Q\} \ge (p-h)/p\}$. Q.E.D.

To quantify the computational impact of Lemma 2, we compute the size of the ordering search space in the perfect-information dynamic program (10) both before and after applying the lemma. Let $\mathcal{G}, \mathcal{Y}, \mathcal{P}$, and $\mathcal{A}$ denote the discretized goodwill, inventory, price, and advertising grids, respectively. Absent structural restrictions, the order-up-to decision is searched over $\{0, \ldots, \bar{N}\}$ for each $(g, y, p, a) \in \mathcal{G} \times \mathcal{Y} \times \mathcal{P} \times \mathcal{A}$, yielding a total of $|\mathcal{G}|\,|\mathcal{Y}|\,|\mathcal{P}|\,|\mathcal{A}|\,(\bar{N}+1)$ baseline evaluations per period.[1] Lemma 2 therefore admits the following reduced number of evaluations:

$$\text{reduced evaluations} \;=\; \sum_{(g,y,p,a)} \left(\max\{0, Q_m^* - y\} + 1\right),$$

where Q_m^* is the newsvendor quantity defined in the lemma. The percentage reduction is

$$1 - \frac{\text{reduced evaluations}}{\text{baseline evaluations}},$$

which isolates the computational savings attributable to the structural bound. To illustrate the size of this reduction, consider the following discretization for our numerical studies in Section 7: $\bar{N} = 3$, $|\mathcal{G}| = 4$, $|\mathcal{Y}| = 4$, $|\mathcal{P}| = 5$, and $|\mathcal{A}| = 17$. The baseline formulation evaluates $(\bar{N}+1) = 4$ ordering decisions for each (g, y, p, a), resulting in $|\mathcal{G}|\,|\mathcal{Y}|\,|\mathcal{P}|\,|\mathcal{A}|\,(\bar{N}+1) = 5{,}440$ evaluations per period. In this example, Lemma 2 reduces the number of evaluations by 71% (from 5,440 to 1,578 per period), holding fixed the same state and action grids.

A.3. Additional Results on Managing Goodwill and Inventory

We conclude this section with two additional lemmas. The first lemma provides a threshold over T such that $g_{\max}$ can be reached in at most $n < \log T$ periods.

Lemma 3. *Let* $\lambda(a_{\max}) > (1-\beta) g_{\max}$ *and* $T > \exp\left(\left\lceil \log\left(\frac{\lambda(a_{\max}) - (1-\beta) g_{\max}}{\lambda(a_{\max}) - \lambda(a_{\min})}\right) / \log\beta \right\rceil\right)$, *where* $\lceil\cdot\rceil$ *denotes the ceiling function (i.e., for* $x \in \mathbb{R}$, $\lceil x \rceil$ *is the least integer greater than or equal to* x*). Then, there exists* $n < \log T$ *such that, starting from* $g_0 = g_{\min}$ *and setting* $a_t = a_{\max}$ *for* $t = 1, \ldots, n$, *we have that* $g_n = g_{\max}$.

[1] As the dynamic program is solved by backward induction over T periods, the total computational effort scales linearly with T.

Proof of Lemma 3. Recall that the initial goodwill level is $g_{\min} = \lambda(a_{\min})/(1-\beta)$. Since $a_t = a_{\max}$ for $t = 1, \dots, n$, we deduce from (1) that $g_n = \min\left\{\frac{\beta^n \lambda(a_{\min}) + (1-\beta^n)\lambda(a_{\max})}{(1-\beta)}, g_{\max}\right\}$, which equals $g_{\max}$ if and only if $n \ge \log\left(\left(\frac{\lambda(a_{\max}) - (1-\beta) g_{\max}}{\lambda(a_{\max}) - \lambda(a_{\min})}\right)^{\frac{1}{\log \beta}}\right)$. Because $T > \exp\left(\left\lceil \log\left(\frac{\lambda(a_{\max}) - (1-\beta) g_{\max}}{\lambda(a_{\max}) - \lambda(a_{\min})}\right) / \log \beta \right\rceil\right)$, choosing n as $\left\lceil \log\left(\frac{\lambda(a_{\max}) - (1-\beta) g_{\max}}{\lambda(a_{\max}) - \lambda(a_{\min})}\right) / \log \beta \right\rceil$ ensures that $g_n = g_{\max}$. That is, starting from $g_0 = g_{\min}$ and setting $a_t = a_{\max}$ for $t = 1, \dots, n$, $g_{\max}$ can be reached in at the end of n periods. Q.E.D.

The second lemma provides a threshold over the holding cost h below which the seller finds having zero inventory never optimal when entering a period.

Lemma 4. *Let $h < p_{\min}\big(1 - (1 - \rho(p_{\max}) g_{\min})^{\bar{N}}\big)$. Then, in the dynamic program (10), it is without loss of optimality to restrict attention to $q \in \{1, \dots, \bar{N}\}$ whenever $\hat{y}_{t-1} = 0$ and $\hat{g}_{t-1} \in [g_{\min}, g_{\max}]$.*

Proof of Lemma 4. We prove the lemma by induction on t. For the base step ($t = T$), we show that focusing on $q \in \{1, \dots, \bar{N}\}$ is without loss of optimality when $\hat{y}_{T-1} = 0$. To this end, note that

$$\hat{V}_T^{(j,k)}\big((\hat{g}_{T-1}, 0)\big) = \max_{\substack{a \in [a_{\min}, a_{\max}] \\ p \in [p_{\min}, p_{\max}] \\ q \in \{0, \dots, \bar{N}\}}} \mathbb{E}^{(j,k)}\left[p \min\{D_T, q\} - a - hq\right].$$

At $q = 0$, the right-hand side equals $-a_{\min}$. At $q = 1$, we obtain the following lower bound on the right-hand side:

$$\max_{\substack{a \in [a_{\min}, a_{\max}] \\ p \in [p_{\min}, p_{\max}]}} \left\{p\, \mathbb{P}^{(j,k)}\{D_T \ge 1\} - a - h\right\} \ge p_{\min}\big(1 - (1 - \rho_j(p_{\max}) g_{\min})^{\bar{N}}\big) - a_{\min} - h.$$

To derive this lower bound, we use the fact that D_T follows a binomial distribution with $\bar{N}$ trials and a success probability of $\rho_j(p_T) g_T \ge \rho_j(p_{\max}) g_{\min}$ (because $\rho_j(\cdot)$ is decreasing and $g_T \ge g_{\min}$). Since $h < p_{\min}\big(1 - (1 - \rho(p_{\max}) g_{\min})^{\bar{N}}\big)$, the preceding lower bound strictly exceeds $-a_{\min}$. Thus, $q = 0$ is never optimal when $\hat{y}_{T-1} = 0$. Moreover, a simple change of variable from order quantity to order-up-to level (as in the proof of Proposition 1(ii)) yields $\hat{V}_T^{(j,k)}\big((\hat{g}_{T-1}, 0)\big) = \hat{V}_T^{(j,k)}\big((\hat{g}_{T-1}, 1)\big)$.

For the induction step, assume that $\hat{V}_{t+1}^{(j,k)}\big((\hat{g}_t, 0)\big) = \hat{V}_{t+1}^{(j,k)}\big((\hat{g}_t, 1)\big)$ for some $t < T$. When $\hat{y}_{t-1} = 0$, the right-hand side of (10) at $q = 0$ is $\max_{a \in [a_{\min}, a_{\max}]}\left\{-a + \hat{V}_{t+1}^{(j,k)}\big(\min\{\beta \hat{g}_{t-1} + \lambda_k(a), g_{\max}\}, 0\big)\right\}$. At $q = 1$, we obtain the following lower bound on the right-hand side of (10):

$$\begin{aligned}
&\max_{\substack{a \in [a_{\min}, a_{\max}] \\ p \in [p_{\min}, p_{\max}]}} \left\{p\, \mathbb{P}^{(j,k)}\{D_t \ge 1\} - a - h + \hat{V}_{t+1}^{(j,k)}\big((\min\{\beta \hat{g}_{t-1} + \lambda_k(a), g_{\max}\}, 1)\big)\right\} \\
&\ge p_{\min}\big(1 - (1 - \rho_j(p_{\max}) g_{\min})^{\bar{N}}\big) - h + \max_{a \in [a_{\min}, a_{\max}]} \left\{-a + \hat{V}_{t+1}^{(j,k)}\big((\min\{\beta \hat{g}_{t-1} + \lambda_k(a), g_{\max}\}, 1)\big)\right\} \\
&= p_{\min}\big(1 - (1 - \rho_j(p_{\max}) g_{\min})^{\bar{N}}\big) - h + \max_{a \in [a_{\min}, a_{\max}]} \left\{-a + \hat{V}_{t+1}^{(j,k)}\big((\min\{\beta \hat{g}_{t-1} + \lambda_k(a), g_{\max}\}, 0)\big)\right\},
\end{aligned}$$

where the inequality follows from the binomial distribution of D_t, and the equality follows from the induction hypothesis. As $h < p_{\min}\big(1 - (1 - \rho(p_{\max}) g_{\min})^{\bar{N}}\big)$, the first term is positive, and hence, the profit from $q = 1$ dominates that from $q = 0$. Thus, $q = 0$ is never optimal when $\hat{y}_{t-1} = 0$, and $\hat{V}_t^{(j,k)}\big((\hat{g}_{t-1}, 0)\big) = \hat{V}_t^{(j,k)}\big((\hat{g}_{t-1}, 1)\big)$. By induction, the result holds for $t = 1, \dots, T$. Q.E.D.

Appendix B: Proofs of the Results in Section 4

Proof of Lemma 1. We prove this result by induction. For the base step, consider period $\ell=1$. We know from (1) that $g_{k,1}=\beta g_0+\lambda_k(a_1)$ for $k\in\{0,1\}$, and because $a_1\geq a_{\rm int}$, $\lambda_1(a_1)\geq\lambda_0(a_1)$. Furthermore, we deduce from (17) that $g_{\max}\geq\beta g_0+\lambda_k(a_1)$ for $k\in\{0,1\}$. Therefore,

$$|g_{1,1}-g_{0,1}|=g_{1,1}-g_{0,1}=\lambda_1(a_1)-\lambda_0(a_1)\geq\delta_g.$$

Thus, (12) holds for $\ell=1$. Moreover, as there is positive inventory in period 1, $\mathbb{P}^{(j,k)}\{C_1\geq 1\}=1-(1-g_{k,1})^{\bar N}$. Consequently, when $g_{k,1}\geq 1-(1-\delta_C)^{1/\bar N}$, (14) is satisfied for $\ell=1$. Because a_1 satisfies (16), it follows that $g_{k,1}\geq 1-(1-\delta_C)^{1/\bar N}$, and hence, (14) holds for $\ell=1$.

For the induction step, consider period $\ell>1$, and suppose that $g_{\max}\geq g_{k,\ell-1}$, $g_{1,\ell-1}-g_{0,\ell-1}\geq\delta_g$, and $g_{k,\ell-1}\geq 1-(1-\delta_C)^{1/\bar N}$ for $k\in\{0,1\}$. Using (1), (17), and the fact that $g_{\max}\geq g_{k,\ell-1}$, we deduce that $g_{k,\ell}=\beta g_{k,\ell-1}+\lambda_k(a_\ell)$. Based on this observation, we verify (12) as follows:

$$|g_{1,\ell}-g_{0,\ell}|=\beta(g_{1,\ell-1}-g_{0,\ell-1})+\lambda_1(a_\ell)-\lambda_0(a_\ell)\geq\beta\delta_g+(1-\beta)\delta_g=\delta_g.$$

Finally, because there is positive inventory in each period, the induction hypothesis that $g_{k,\ell-1}\geq 1-(1-\delta_C)^{1/\bar N}$ and (16) jointly imply that (14) holds. Q.E.D.

Proof of Theorem 1. To prove this result, we first derive (19) and then (20).

Step 1: Proof of (19). For this step, we introduce some notation. For $m\in\{1,\ldots,t\}$, define the event $\mathcal{A}_{t,m}\triangleq\left\{\sum_{\ell=1}^t\mathbb{1}_{\{C_\ell\geq 1\}}=m\right\}$ and the random set $\mathcal{L}_t\triangleq\{\ell\in\{1,\ldots,t\}:C_\ell\geq 1\}$. On $\mathcal{A}_{t,m}$, $\{C_\ell,\,\ell=1,\ldots,t\}$ becomes positive in exactly m periods. The set $\mathcal{L}_t$ corresponds to the collection of periods in which $\{C_\ell,\,\ell=1,\ldots,t\}$ becomes positive. Based on these definitions, for any $L\subseteq\{1,\ldots,t\}$ with cardinality m, let $\mathcal{A}^L_{m,t}=\mathcal{A}_{m,t}\cap\{\mathcal{L}_t=L\}$. We first rewrite $\mathbb{E}^{(0,k)}_{\mathbf{\Pi}}[b_t]$ by conditioning on the events $\{\mathcal{A}^L_{t,m}\}$ as follows:

$$\mathbb{E}^{(0,k)}_{\mathbf{\Pi}}[b_t]=\sum_{m=0}^{t}\sum_{L\subseteq\{1,\ldots,t\},\,|L|=m}\mathbb{E}^{(0,k)}_{\mathbf{\Pi}}[b_t\,|\,\mathcal{A}^L_{t,m}]\,\mathbb{P}^{(0,k)}_{\mathbf{\Pi}}\{\mathcal{A}^L_{t,m}\},\tag{B.1}$$

where $|L|$ denotes the cardinality of L. To analyze the first term on inside the sum in (B.1), we use the following lemma, whose proof is deferred to the end of this section.

Lemma 5. *Let $\mathbf{\Pi}$ be a $(\boldsymbol{\delta},t)$-discriminative policy satisfying* (18). *Then,*

$$\mathbb{E}^{(0,k)}_{\mathbf{\Pi}}[b_t\,|\,\mathcal{A}^L_{t,m}]\leq\max\left\{2,\frac{b_0}{1-b_0}\right\}\exp\left(-\frac{\delta_\rho^4 m}{2\bar N^2\gamma}\right)$$

for $L\subseteq\{1,\ldots,t\}$ with cardinality $m\in\{0,\ldots,t\}$.

We now combine (B.1) with Lemma 5. Because $\mathbf{\Pi}$ satisfies (14), $\mathbb{P}^{(0,k)}_{\mathbf{\Pi}}\{\mathcal{A}^L_{t,m}\}\leq(1-\delta_C)^{t-m}$. Thus, (B.1) and Lemma 5 imply that

$$\mathbb{E}^{(0,k)}_{\mathbf{\Pi}}[b_t]\leq\sum_{m=0}^{t}\binom{t}{m}(1-\delta_C)^{t-m}\max\left\{2,\frac{b_0}{1-b_0}\right\}\exp\left(-\frac{\delta_\rho^4 m}{2\bar N^2\gamma}\right)$$

$$= \max\left\{2, \frac{b_0}{1-b_0}\right\}\left[1-\delta_C+\exp\left(-\frac{\delta_\rho^4}{2\bar{N}^2\gamma}\right)\right]^t,$$

where the inequality follows because there are $\binom{t}{m}$ subsets of $\{1,\ldots,t\}$ with cardinality m, and the equality follows from binomial expansion. Note here that $1-\delta_C+\exp\left(-\frac{\delta_\rho^4}{2\bar{N}^2\gamma}\right)$ is positive and smaller than 1 because (18) implies that $\delta_C > \exp\left(-\delta_\rho^4/(2\bar{N}^2\gamma)\right)$. This proves that

$$\mathbb{E}_{\mathbf{\Pi}}^{(0,k)}[b_t] \le \mu\exp(-\xi t),$$

where

$$\mu \triangleq \max\left\{2, \frac{b_0}{1-b_0}\right\} \text{ and } \xi \triangleq -\log\left(1-\delta_C+\exp\left(-\frac{\delta_\rho^4}{2\bar{N}^2\gamma}\right)\right). \tag{B.2}$$

We repeat the same analysis to prove that $\mathbb{E}_{\mathbf{\Pi}}^{(1,k)}[1-b_t] \le \mu\exp(-\xi t)$.

Step 2: Proof of (20). Without loss of generality, we focus on the first inequality $\mathbb{E}_{\mathbf{\Pi}}^{(j,0)}[\tilde{b}_t] \le \tilde{\mu}\exp(-\tilde{\xi}t)$, and the analysis that follows can be repeated verbatim for $\mathbb{E}_{\mathbf{\Pi}}^{(j,1)}[1-\tilde{b}_t]$. Using Bayes' rule repeatedly, we first obtain the following expression for $\tilde{b}_t$:

$$\tilde{b}_t = \frac{\tilde{b}_0\prod_{\ell=1}^t l_{\bar{N}}(g_{1,\ell},N_\ell)}{\tilde{b}_0\prod_{\ell=1}^t l_{\bar{N}}(g_{1,\ell},N_\ell) + (1-\tilde{b}_0)\prod_{\ell=1}^t l_{\bar{N}}(g_{0,\ell},N_\ell)}.$$

Consequently,

$$\mathbb{E}_{\mathbf{\Pi}}^{(j,0)}[\tilde{b}_t] = \mathbb{E}_{\mathbf{\Pi}}^{(j,0)}\left[\frac{1}{1+\frac{1-\tilde{b}_0}{\tilde{b}_0}\exp\left(\sum_{\ell=1}^t \log\frac{l_{\bar{N}}(g_{0,\ell},N_\ell)}{l_{\bar{N}}(g_{1,\ell},N_\ell)}\right)}\right] \tag{B.3}$$

Using the definition of $l_{\bar{N}}(\cdot,\cdot)$, we first simplify the argument of the exponential function in the preceding equation and then derive a lower bound on it as follows:

$$\begin{aligned}\sum_{\ell=1}^t \log\frac{l_{\bar{N}}(g_{0,\ell},N_\ell)}{l_{\bar{N}}(g_{1,\ell},N_\ell)} &= \sum_{\ell=1}^t\left[(N_\ell-\bar{N}g_{0,\ell})\log\frac{g_{0,\ell}(1-g_{1,\ell})}{g_{1,\ell}(1-g_{0,\ell})}+\bar{N}\left(\log\frac{1-g_{0,\ell}}{1-g_{1,\ell}}+g_{0,\ell}\log\frac{g_{0,\ell}(1-g_{1,\ell})}{g_{1,\ell}(1-g_{0,\ell})}\right)\right]\\ &\ge \sum_{\ell=1}^t\left((N_\ell-\bar{N}g_{0,\ell})\log\frac{g_{0,\ell}(1-g_{1,\ell})}{g_{1,\ell}(1-g_{0,\ell})}+2\bar{N}\delta_g^2\right),\end{aligned} \tag{B.4}$$

where the inequality follows from (12) and Lemma A.2 of Harrison et al. (2012). Letting $\tilde{\zeta}_t \triangleq \sum_{\ell=1}^t (N_\ell-\bar{N}g_{0,\ell})\log\frac{g_{0,\ell}(1-g_{1,\ell})}{g_{1,\ell}(1-g_{0,\ell})}$, we deduce from (B.3) and (B.4) that

$$\mathbb{E}_{\mathbf{\Pi}}^{(j,0)}[\tilde{b}_t] \le \frac{1}{1+\frac{1-\tilde{b}_0}{\tilde{b}_0}\exp(\bar{N}\delta_g^2 t)} + \mathbb{P}_{\mathbf{\Pi}}^{(j,0)}\left\{|\tilde{\zeta}_t| \ge \bar{N}\delta_g^2 t\right\}. \tag{B.5}$$

We now derive an upper bound on $\mathbb{P}_{\mathbf{\Pi}}^{(j,0)}\left\{|\tilde{\zeta}_t| \ge \bar{N}\delta_g^2 t\right\}$. Let $\tilde{\zeta}_0 = 0$, and $\tilde{\mathcal{F}}_t$ be the filtration generated by $\{N_\ell,\, \ell=1,\ldots,t\}$. Note that

$$\begin{aligned}\mathbb{E}_{\mathbf{\Pi}}^{(j,0)}\left[\tilde{\zeta}_t \,|\, \tilde{\mathcal{F}}_{t-1}\right] &= \mathbb{E}_{\mathbf{\Pi}}^{(j,0)}\left[\sum_{\ell=1}^t (N_\ell-\bar{N}g_{0,\ell})\log\frac{g_{0,\ell}(1-g_{1,\ell})}{g_{1,\ell}(1-g_{0,\ell})}\,\middle|\, \tilde{\mathcal{F}}_{t-1}\right]\\ &= \tilde{\zeta}_{t-1} + \mathbb{E}_{\mathbf{\Pi}}^{(j,0)}\left[(N_t-\bar{N}g_{0,t})\log\frac{g_{0,t}(1-g_{1,t})}{g_{1,t}(1-g_{0,t})}\,\middle|\, \tilde{\mathcal{F}}_{t-1}\right] = \tilde{\zeta}_{t-1},\end{aligned}$$

where the last equality follows because, under hypothesis $(j,0)$, N_t is a binomial random variable with parameters $\bar{N}$ and $g_{0,t}$. Thus, $\{\tilde{\zeta}_t, t=1,2,\ldots\}$ is a martingale adapted to $\tilde{\mathcal{F}}_t$. Moreover, the martingale differences are bounded by $\bar{N}\sqrt{\tilde{\gamma}}$, where $\tilde{\gamma}=[\log(g_{\max}(1-g_{\min}))-\log(g_{\min}(1-g_{\max}))]^2$. Consequently, Azuma's inequality (Ross 1996, p. 307) implies that

$$\mathbb{P}_{\mathbf{\Pi}}^{(j,0)}\left\{|\tilde{\zeta}_t|\ge \bar{N}\delta_g^2 t\right\}\le 2\exp\left(-\frac{\bar{N}^2\delta_g^4 t^2}{2t\bar{N}^2\tilde{\gamma}}\right)=2\exp\left(-\frac{\delta_g^4 t}{2\tilde{\gamma}}\right). \tag{B.6}$$

Combining (B.5) and (B.6), we deduce that

$$\mathbb{E}_{\mathbf{\Pi}}^{(j,0)}[\tilde{b}_t]\le \frac{1}{1+\frac{1-\tilde{b}_0}{\tilde{b}_0}\exp(\bar{N}\delta_g^2 t)}+\mathbb{P}_{\mathbf{\Pi}}^{(j,0)}\left\{|\tilde{\zeta}_t|\ge \bar{N}\delta_g^2 t\right\}\le \frac{\tilde{b}_0}{1-\tilde{b}_0}\exp(-\bar{N}\delta_g^2 t)+2\exp\left(-\frac{\delta_g^4 t}{2\tilde{\gamma}}\right)\le \tilde{\mu}\exp(-\tilde{\xi}t),$$

where

$$\tilde{\mu}\triangleq \max\left\{2,\frac{\tilde{b}_0}{1-\tilde{b}_0}\right\}\ \text{and}\ \tilde{\xi}\triangleq \min\left\{\bar{N}\delta_g^2,\frac{\delta_g^4}{2\tilde{\gamma}}\right\}. \tag{B.7}$$

This concludes the proof. Q.E.D.

Proof of Lemma 5. We prove this result in two steps.

Step 1. We first derive an upper bound on $\mathbb{E}_{\mathbf{\Pi}}^{(0,k)}[b_t\,|\,\mathcal{A}_{t,m}^L]$. By Bayes' rule, we deduce that

$$b_t=\frac{b_0\prod_{\ell\in\mathcal{L}_t}\prod_{i=1}^{C_\ell} l_1(\rho_1(p_\ell),S_{\ell,i})}{b_0\prod_{\ell\in\mathcal{L}_t}\prod_{i=1}^{C_\ell} l_1(\rho_1(p_\ell),S_{\ell,i})+(1-b_0)\prod_{\ell\in\mathcal{L}_t}\prod_{i=1}^{C_\ell} l_1(\rho_0(p_\ell),S_{\ell,i})}=\left[1+\frac{1-b_0}{b_0}\prod_{\ell\in\mathcal{L}_t}\prod_{i=1}^{C_\ell}\frac{l_1(\rho_0(p_\ell),S_{\ell,i})}{l_1(\rho_1(p_\ell),S_{\ell,i})}\right]^{-1}.$$

Therefore,

$$\mathbb{E}_{\mathbf{\Pi}}^{(0,k)}[b_t\,|\,\mathcal{A}_{t,m}^L]=\mathbb{E}_{\mathbf{\Pi}}^{(0,k)}\left[\left[1+\frac{1-b_0}{b_0}\exp\left(\sum_{\ell\in\mathcal{L}_t}\sum_{i=1}^{C_\ell}\log\frac{l_1(\rho_0(p_\ell),S_{\ell,i})}{l_1(\rho_1(p_\ell),S_{\ell,i})}\right)\right]^{-1}\,\middle|\,\mathcal{A}_{t,m}^L\right].$$

We next focus on the argument of the preceding exponential function. On $\mathcal{A}_{t,m}^L$, we have that

$$\begin{aligned}\sum_{\ell\in\mathcal{L}_t}\sum_{i=1}^{C_\ell}\log\frac{l_1(\rho_0(p_\ell),S_{\ell,i})}{l_1(\rho_1(p_\ell),S_{\ell,i})}&=\sum_{\ell\in\mathcal{L}_t}\sum_{i=1}^{C_\ell}\log\frac{(\rho_0(p_\ell))^{S_{\ell,i}}(1-\rho_0(p_\ell))^{1-S_{\ell,i}}}{(\rho_1(p_\ell))^{S_{\ell,i}}(1-\rho_1(p_\ell))^{1-S_{\ell,i}}}\\&\ge\sum_{\ell\in\mathcal{L}_t}\left[2\delta_\rho^2+\sum_{i=1}^{C_\ell}[S_{\ell,i}-\rho_0(p_\ell)]\log\frac{\rho_0(p_\ell)(1-\rho_1(p_\ell))}{\rho_1(p_\ell)(1-\rho_0(p_\ell))}\right],\end{aligned}$$

where the equality follows because there is no sales observations when either there is a stock out or no customer arrives (i.e., $C_\ell=0$), and the inequality follows from (13) and Lemma A.2 of Harrison et al. (2012). On $\mathcal{A}_{t,m}^L$, $\sum_{\ell=1}^t\mathbb{1}_{\{C_\ell\ge1\}}=m$; hence, we obtain the following upper bound:

$$\mathbb{E}_{\mathbf{\Pi}}^{(0,k)}[b_t\,|\,\mathcal{A}_{t,m}^L]\le\mathbb{E}_{\mathbf{\Pi}}^{(0,k)}\left[\left[1+\frac{1-b_0}{b_0}\exp\left(2\delta_\rho^2 m+\sum_{\ell\in\mathcal{L}_t}\sum_{i=1}^{C_\ell}[S_{\ell,i}-\rho_0(p_\ell)]\log\frac{\rho_0(p_\ell)(1-\rho_1(p_\ell))}{\rho_1(p_\ell)(1-\rho_0(p_\ell))}\right)\right]^{-1}\,\middle|\,\mathcal{A}_{t,m}^L\right].$$

Thus, considering the cases where the absolute value of the sum inside the preceding exponential function is larger or smaller than $\delta_\rho^2 m$, we deduce that

$$\mathbb{E}_{\mathbf{\Pi}}^{(0,k)}[b_t\,|\,\mathcal{A}_{t,m}^L]\le\frac{1}{1+\frac{1-b_0}{b_0}\exp(\delta_\rho^2 m)}+\mathbb{P}_{\mathbf{\Pi}}^{(0,k)}\left\{\left|\sum_{\ell\in\mathcal{L}_t}\sum_{i=1}^{C_\ell}[S_{\ell,i}-\rho_0(p_\ell)]\log\frac{\rho_0(p_\ell)(1-\rho_1(p_\ell))}{\rho_1(p_\ell)(1-\rho_0(p_\ell))}\right|\ge\delta_\rho^2 m\,\middle|\,\mathcal{A}_{t,m}^L\right\}. \tag{B.8}$$

Step 2. In this step, we derive an upper bound on the probability in (B.8), and then we use (B.8) to obtain a bound on $\mathbb{E}_{\mathbf{\Pi}}^{(0,k)}[b_t \,|\, \mathcal{A}_{t,m}^L]$. Now, let $\vartheta = \sum_{\ell\in\mathcal{L}_t}\sum_{i=1}^{C_\ell}[S_{\ell,i}-\rho_0(p_\ell)]\log\frac{\rho_0(p_\ell)(1-\rho_1(p_\ell))}{\rho_1(p_\ell)(1-\rho_0(p_\ell))}$. To derive an upper bound on $\mathbb{P}_{\mathbf{\Pi}}^{(0,k)}\{|\vartheta|\geq\delta_\rho^2 m \,|\, \mathcal{A}_{t,m}^L\}$, we note that $\mathbb{P}_{\mathbf{\Pi}}^{(0,k)}\{|\vartheta|\geq\delta_\rho^2 m \,|\, \mathcal{A}_{t,m}^L\} \leq \mathbb{P}_{\mathbf{\Pi}}^{(0,k)}\{\vartheta\geq\delta_\rho^2 m \,|\, \mathcal{A}_{t,m}^L\}+\mathbb{P}_{\mathbf{\Pi}}^{(0,k)}\{-\vartheta\geq\delta_\rho^2 m \,|\, \mathcal{A}_{t,m}^L\}$. We first focus on $\mathbb{P}_{\mathbf{\Pi}}^{(0,k)}\{\vartheta\geq\delta_\rho^2 m \,|\, \mathcal{A}_{t,m}^L\}$. Letting $\epsilon>0$ and applying Markov's inequality, we obtain the following:

$$\mathbb{P}_{\mathbf{\Pi}}^{(0,k)}\{\vartheta\geq\delta_\rho^2 m \,|\, \mathcal{A}_{t,m}^L\} \leq \exp\left(-\epsilon\delta_\rho^2 m\right)\prod_{\ell\in\mathcal{L}_t}\mathbb{E}_{\mathbf{\Pi}}^{(0,k)}\left[\exp(\epsilon\upsilon_\ell)\,\big|\,\mathcal{A}_{t,m}^L\right], \tag{B.9}$$

where $\upsilon_\ell=\sum_{i=1}^{C_\ell}[S_{\ell,i}-\rho_0(p_\ell)]\log\frac{\rho_0(p_\ell)(1-\rho_1(p_\ell))}{\rho_1(p_\ell)(1-\rho_0(p_\ell))}$ so that $\vartheta=\sum_{\ell\in\mathcal{L}_t}\upsilon_\ell$. Note that $-\bar{N}\sqrt{\gamma}\leq\upsilon_\ell\leq\bar{N}\sqrt{\gamma}$. Moreover, under hypothesis $(0,k)$, $S_{\ell,i}$ is a Bernoulli random variable with success probability $\rho_0(p_\ell)$ for all $i\leq C_\ell$ (i.e., C_ℓ is a stopping time); thus, we deduce from Wald's equation (Ross 1996, p. 300) that $\mathbb{E}_{\mathbf{\Pi}}^{(0,k)}[\upsilon_\ell \,|\, \mathcal{A}_{t,m}^L]=0$. Hence, Hoeffding's lemma (Massart 2007, p. 21) implies that $\mathbb{E}_{\mathbf{\Pi}}^{(0,k)}[\exp(\epsilon\upsilon_\ell) \,|\, \mathcal{A}_{t,m}^L]\leq\exp\left(\frac{\epsilon^2\bar{N}^2\gamma}{2}\right)$. Using this bound in (B.9) and optimizing over ϵ, we deduce that $\mathbb{P}_{\mathbf{\Pi}}^{(0,k)}\{\vartheta\geq\delta_\rho^2 m \,|\, \mathcal{A}_{t,m}^L\}\leq\min_{\epsilon>0}\left\{\exp\left(-\epsilon\delta_\rho^2 m\right)\exp\left(\frac{m\epsilon^2\bar{N}^2\gamma}{2}\right)\right\}\leq\exp\left(-\frac{\delta_\rho^4 m}{2\bar{N}^2\gamma}\right)$. Repeating these calculations leads to $\mathbb{P}_{\mathbf{\Pi}}^{(0,k)}\left\{-\vartheta\geq\delta_\rho^2 m \,|\, \mathcal{A}_{t,m}^L\right\}\leq\exp\left(-\frac{\delta_\rho^4 m}{2\bar{N}^2\gamma}\right)$. Therefore, $\mathbb{P}_{\mathbf{\Pi}}^{(0,k)}\left\{|\vartheta|\geq\delta_\rho^2 m \,|\, \mathcal{A}_{t,m}^L\right\}\leq 2\exp\left(-\frac{\delta_\rho^4 m}{2\bar{N}^2\gamma}\right)$. Using this upper bound in (B.8), we obtain that

$$\begin{aligned}\mathbb{E}_{\mathbf{\Pi}}^{(0,k)}[b_t \,|\, \mathcal{A}_{t,m}^L] &\leq \frac{1}{1+\frac{1-b_0}{b_0}\exp(\delta_\rho^2 m)}+2\exp\left(-\frac{\delta_\rho^4 m}{2\bar{N}^2\gamma}\right)\leq\frac{b_0}{1-b_0}\exp(-\delta_\rho^2 m)+2\exp\left(-\frac{\delta_\rho^4 m}{2\bar{N}^2\gamma}\right)\\ &\leq\max\left\{2,\frac{b_0}{1-b_0}\right\}\exp\left(-\frac{\delta_\rho^4 m}{2\bar{N}^2\gamma}\right).\quad \text{Q.E.D.}\end{aligned}$$

Appendix C: Proof of Theorem 2

We begin with introducing some notation. Let $\mathbf{\Pi}^{\mathsf{3P}}$ be a three-phase policy defined as follows. In the first τ_1 periods, $\mathbf{\Pi}^{\mathsf{3P}}$ uses a discriminative policy. For $t=\tau_1+1,\ldots,\tau_1+\tau_2$, $\mathbf{\Pi}^{\mathsf{3P}}$ chooses $a_t=a_{\max}$, $p_t=p_{\min}$, and $q_t=0$. In the remaining periods, $\mathbf{\Pi}^{\mathsf{3P}}$ uses the optimal perfect information policy for the most likely hypothesis according to the seller's belief at the end of period τ_1. Note that, under hypothesis (j,k) and policy $\tilde{\mathbf{\Pi}}$, the seller's expected profit from period m_1 to m_2 is

$$J^{(j,k)}(m_1,m_2 \,|\, \tilde{\mathbf{\Pi}}) \triangleq \mathbb{E}_{\tilde{\mathbf{\Pi}}}^{(j,k)}\left[\sum_{t=m_1}^{m_2}\left(\tilde{p}_t\min\{D_t,\tilde{y}_{t-1}+\tilde{q}_t\}-\tilde{a}_t-h(\tilde{y}_{t-1}+\tilde{q}_t)\right)\right], \tag{C.1}$$

where $\{\tilde{p}_t\}$, $\{\tilde{a}_t\}$, and $\{\tilde{q}_t\}$ are the decisions made by $\tilde{\mathbf{\Pi}}$; $\{\tilde{y}_{t-1}\}$ and $\{\tilde{g}_{t-1}\}$ are the induced inventory and goodwill levels, respectively; and D_t is a random variable following a binomial distribution with $\bar{N}$ trials and a success probability of $\rho_j(\tilde{p}_t)\min\{g_{\max},\beta\tilde{g}_{t-1}+\lambda_k(\tilde{a}_t)\}$.

Now, fix hypothesis (j,k). Based on (C.1), we decompose $\mathsf{R}^{(j,k)}\left(\mathbf{\Pi}^{\mathsf{3P}}\right)$ as follows:

$$\begin{aligned}\mathsf{R}^{(j,k)}\left(\mathbf{\Pi}^{\mathsf{3P}}\right) &= J^{(j,k)}(1,\tau_1 \,|\, \hat{\mathbf{\Pi}})-J^{(j,k)}(1,\tau_1 \,|\, \mathbf{\Pi}^{\mathsf{3P}})\\ &\quad+J^{(j,k)}(\tau_1+1,\tau_1+\tau_2 \,|\, \hat{\mathbf{\Pi}})-J^{(j,k)}(\tau_1+1,\tau_1+\tau_2 \,|\, \mathbf{\Pi}^{\mathsf{3P}})\\ &\quad+J^{(j,k)}(\tau_1+\tau_2+1,T \,|\, \hat{\mathbf{\Pi}})-J^{(j,k)}(\tau_1+\tau_2+1,T \,|\, \mathbf{\Pi}^{\mathsf{3P}}).\end{aligned} \tag{C.2}$$

To analyze the terms on right-hand side of (C.2), we state the following lemma, the proof of which is deferred to the end of this section.

LEMMA 6. *There exist constants $\Psi_1 > 0$, $\Psi_2 > 0$, and $\Psi_3 > 1$ such that, under any hypothesis (j,k),*

(i) $J^{(j,k)}(1,\tau_1 \,|\, \hat{\mathbf{\Pi}}) - J^{(j,k)}(1,\tau_1 \,|\, \mathbf{\Pi}^{\mathsf{3P}}) \le \Psi_1 \tau_1$,

(ii) $J^{(j,k)}(\tau_1+1,\tau_1+\tau_2 \,|\, \hat{\mathbf{\Pi}}) - J^{(j,k)}(\tau_1+1,\tau_1+\tau_2 \,|\, \mathbf{\Pi}^{\mathsf{3P}}) \le \Psi_1 \tau_2$,

(iii) $J^{(j,k)}(\tau_1+\tau_2+1,T \,|\, \hat{\mathbf{\Pi}}) - J^{(j,k)}(\tau_1+\tau_2+1,T \,|\, \mathbf{\Pi}^{\mathsf{3P}}) \le \Psi_2 \Psi_3^{-\tau_1} T$.

REMARK 7. In Lemma 6, we have $\Psi_1 \triangleq \bar{N} p_{\max}$, $\Psi_2 \triangleq 4\Psi_1 \max\{\mu,\tilde{\mu}\}$, and $\Psi_3 \triangleq \exp(\min\{\xi,\tilde{\xi}\})$, where μ, $\tilde{\mu}$, ξ, and $\tilde{\xi}$ are as in (B.2) and (B.7).

Following Lemma 6, we bound $\mathsf{R}^{(j,k)}\left(\mathbf{\Pi}^{\mathsf{3P}}\right)$ as follows:

$$\mathsf{R}^{(j,k)}\left(\mathbf{\Pi}^{\mathsf{3P}}\right) \le \Psi_1\tau_1 + \Psi_1\tau_2 + \Psi_2\Psi_3^{-\tau_1}T. \tag{C.3}$$

Choose τ_1 equal to the parameter τ of $\mathbf{\Pi}^{\mathsf{JL}}$, and τ_2 equal to n, which is the number of periods in which the maximum possible goodwill $g_{\max}$ would be reached with certainty. By definition, $\tau_1 = \tau = (\log T)/(\log \Psi_3) = \log_{\Psi_3} T$. Then, (C.3) implies that

$$\mathsf{R}^{(j,k)}\left(\mathbf{\Pi}^{\mathsf{3P}}\right) \le \frac{\Psi_1}{\log \Psi_3}\log T + \Psi_1 n + \Psi_2 \Psi_3^{-\log_{\Psi_3} T} T \le 3\max\left\{\frac{\Psi_1}{\log\Psi_3}, \Psi_1, \Psi_2\right\}\log T,$$

where the last inequality follows because $n < \log T$ and $1 \le \log T$. Thus, $\mathsf{R}^{(j,k)}\left(\mathbf{\Pi}^{\mathsf{3P}}\right) \le \mu^{\mathsf{3P}} \log T$, where $\mu^{\mathsf{3P}} = 3\max\left\{\frac{\Psi_1}{\log\Psi_3}, \Psi_1, \Psi_2\right\}$. That is, $\mathbf{\Pi}^{\mathsf{3P}}$ achieves a T-period regret at most of order $\log T$. We obtain the same conclusion for $\mathbf{\Pi}^{\mathsf{JL}}$ because implementing the adjustment phase and then perfect information policy cannot outperform directly implementing the perfect information policy. To see this, first note that

$$\mathsf{R}^{(j,k)}\left(\mathbf{\Pi}^{\mathsf{JL}}\right) = J^{(j,k)}(1,\tau \,|\, \hat{\mathbf{\Pi}}) - J^{(j,k)}(1,\tau \,|\, \mathbf{\Pi}^{\mathsf{JL}}) + J^{(j,k)}(\tau+1,T \,|\, \hat{\mathbf{\Pi}}) - J^{(j,k)}(\tau+1,T \,|\, \mathbf{\Pi}^{\mathsf{JL}}). \tag{C.4}$$

Observe that both $\mathbf{\Pi}^{\mathsf{JL}}$ and $\mathbf{\Pi}^{\mathsf{3P}}$ employ a discriminative policy in the first $\tau_1 = \tau$ periods. Thus, Lemma 6(i) holds with $\mathbf{\Pi}^{\mathsf{3P}}$ and τ_1 replaced by $\mathbf{\Pi}^{\mathsf{JL}}$ and τ, respectively. That is,

$$J^{(j,k)}(1,\tau \,|\, \hat{\mathbf{\Pi}}) - J^{(j,k)}(1,\tau \,|\, \mathbf{\Pi}^{\mathsf{JL}}) \le \Psi_1 \tau. \tag{C.5}$$

Note also that both $\mathbf{\Pi}^{\mathsf{JL}}$ and $\mathbf{\Pi}^{\mathsf{3P}}$ choose which perfect information policy to employ using the same belief, namely, the belief at the end of period τ. For $j',k' \in \{0,1\}$, let $\mathcal{E}^{(j',k')}$ denote the event that hypothesis (j',k') is selected. Then, $1 - \mathbb{P}^{(j,k)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{\mathcal{E}^{(j,k)}\} = 1 - \mathbb{P}^{(j,k)}_{\mathbf{\Pi}^{\mathsf{JL}}}\{\mathcal{E}^{(j,k)}\}$. Consequently,

$$\begin{aligned}
J^{(j,k)}(\tau+1,T \,|\, \hat{\mathbf{\Pi}}) - J^{(j,k)}(\tau+1,T \,|\, \mathbf{\Pi}^{\mathsf{JL}}) &\le J^{(j,k)}(\tau+1,T \,|\, \hat{\mathbf{\Pi}}) - \mathbb{E}^{(j,k)}_{\hat{\mathbf{\Pi}}}\left[\hat{V}^{(j,k)}_{\tau+1}\left((g^{\mathsf{JL}}_{\tau+1}, y^{\mathsf{JL}}_{\tau+1})\right)\right] + \Psi_2\Psi_3^{-\tau}T \\
&= J^{(j,k)}(\tau+1,T \,|\, \hat{\mathbf{\Pi}}) - \mathbb{E}^{(j,k)}_{\hat{\mathbf{\Pi}}}\left[\hat{V}^{(j,k)}_{\tau+1}\left((g^{\mathsf{3P}}_{\tau+1}, y^{\mathsf{3P}}_{\tau+1})\right)\right] + \Psi_2\Psi_3^{-\tau}T \\
&\le J^{(j,k)}(\tau+1,T \,|\, \hat{\mathbf{\Pi}}) - J^{(j,k)}(\tau+1,T \,|\, \mathbf{\Pi}^{\mathsf{3P}}) + \Psi_2\Psi_3^{-\tau}T \\
&\le \Psi_1\tau_2 + 2\Psi_2\Psi_3^{-\tau}T.
\end{aligned} \tag{C.6}$$

In the preceding calculations, $g^{\mathsf{JL}}_{\tau+1}$ and $y^{\mathsf{JL}}_{\tau+1}$ denote the inventory and goodwill levels, respectively, under $\mathbf{\Pi}^{\mathsf{JL}}$ at the end of period $\tau+1$. Similarly, $g^{\mathsf{3P}}_{\tau+1}$ and $y^{\mathsf{3P}}_{\tau+1}$ denote the inventory and goodwill levels, respectively, at under $\mathbf{\Pi}^{\mathsf{3P}}$ the end of period $\tau+1$. The first inequality follows by repeating the arguments used to derive Lemma 6(iii). The equality follows because $g^{\mathsf{JL}}_{\tau+1} = g^{\mathsf{3P}}_{\tau+1}$ and $y^{\mathsf{JL}}_{\tau+1} = y^{\mathsf{3P}}_{\tau+1}$. The second inequality follows because the expected profit of implementing the perfect information policy right after τ is an upper bound for the expected profit collected by $\mathbf{\Pi}^{\mathsf{3P}}$ in the remaining $T-\tau$ periods. The last inequality follows by Lemma 6. Using (C.4)-(C.6) and recalling that $\tau_1 = \tau = \log_{\Psi_3} T$ and $\tau_2 = n$, we conclude that

$$\mathsf{R}^{(j,k)}\left(\mathbf{\Pi}^{\mathsf{JL}}\right) \le \frac{\Psi_1}{\log\Psi_3}\log T + \Psi_1 n + 2\Psi_2\Psi_3^{-\log_{\Psi_3}T}T \le 3\max\left\{\frac{\Psi_1}{\log\Psi_3}, \Psi_1, 2\Psi_2\right\}\log T.$$

As a result, letting

$$\begin{aligned}\mu^{\mathsf{JL}} &\triangleq 3\max\left\{\frac{\Psi_1}{\log\Psi_3}, \Psi_1, 2\Psi_2\right\}\\ &= 3\Psi_1\max\left\{\frac{1}{-\log\left(1-\delta_C+\exp\left(-\frac{\delta_\rho^4}{2\bar{N}^2\gamma}\right)\right)}, \frac{1}{\min\left\{\bar{N}\delta_g^2, \frac{\delta_g^4}{2\tilde{\gamma}}\right\}}, 8\max\{\mu,\tilde{\mu}\}\right\},\end{aligned} \tag{C.7}$$

we conclude that $\mathsf{R}^{(j,k)}\left(\mathbf{\Pi}^{\mathsf{JL}}\right) \le \mu^{\mathsf{JL}}\log T$. Q.E.D.

Proof of Lemma 6. Let (j,k) be the underlying hypothesis. Parts (i) and (ii) of the lemma follow because the expected profit per period is bounded from above by $\Psi_1 = \bar{N}p_{\max}$.

We prove part (iii) of the lemma as follows. Recall that $\mathbf{\Pi}^{\mathsf{3P}}$ selects the most likely hypothesis at the end of period τ_1. For $j',k' \in \{0,1\}$, let $\mathcal{E}^{(j',k')}$ denote the event that hypothesis (j',k') is selected. If hypothesis (j,k) is not selected, the difference $J^{(j,k)}(\tau_1+\tau_2+1, T \mid \hat{\mathbf{\Pi}}) - J^{(j,k)}(\tau_1+\tau_2+1, T \mid \mathbf{\Pi}^{\mathsf{3P}})$ is smaller than $\Psi_1 T$ because the expected profit per period is bounded from above by $\Psi_1 = \bar{N}p_{\max}$. If hypothesis (j,k) is selected, the expected profit acquired by $\mathbf{\Pi}^{\mathsf{3P}}$ between periods $\tau_1+\tau_2+1$ and T is $\hat{V}_{\tau_1+\tau_2+1}\big((g_{\max}, y^{\mathsf{3P}}_{\tau_1+\tau_2+1})\big)$, where $y^{\mathsf{3P}}_{\tau_1+\tau_2+1}$ denotes the inventory level at the end of period $\tau_1+\tau_2+1$ under $\mathbf{\Pi}^{\mathsf{3P}}$. Therefore, we bound the aforementioned difference as follows:

$$\begin{aligned}&J^{(j,k)}(\tau_1+\tau_2+1, T \mid \hat{\mathbf{\Pi}}) - J^{(j,k)}(\tau_1+\tau_2+1, T \mid \mathbf{\Pi}^{\mathsf{3P}})\\ &\quad\le \left[J^{(j,k)}(\tau_1+\tau_2+1, T \mid \hat{\mathbf{\Pi}}) - \hat{V}^{(j,k)}_{\tau_1+\tau_2+1}\big((g_{\max}, y^{\mathsf{3P}}_{\tau_1+\tau_2+1})\big)\right]\mathbb{P}^{(j,k)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{\mathcal{E}^{(j,k)}\} + \Psi_1 T(1-\mathbb{P}^{(j,k)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{\mathcal{E}^{(j,k)}\}).\end{aligned} \tag{C.8}$$

We first analyze the probability $1-\mathbb{P}^{(j,k)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{\mathcal{E}^{(j,k)}\}$ at the right-hand side of (C.8). Without loss of generality, consider the case where $j=k=0$. Therefore, $1-\mathbb{P}^{(j,k)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{\mathcal{E}^{(j,k)}\} = 1-\mathbb{P}^{(0,0)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{\mathcal{E}^{(0,0)}\}$ is the probability that the true hypothesis $(0,0)$ is not selected. This probability is bounded as follows:

$$\begin{aligned}1-\mathbb{P}^{(j,k)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{\mathcal{E}^{(j,k)}\} = 1-\mathbb{P}^{(0,0)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{b^{\mathsf{3P}}_{\tau_1}<1/2\}\mathbb{P}^{(0,0)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{\tilde{b}^{\mathsf{3P}}_{\tau_1}<1/2\} &= 1-[1-\mathbb{P}^{(0,0)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{b^{\mathsf{3P}}_{\tau_1}\ge 1/2\}][1-\mathbb{P}^{(0,0)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{\tilde{b}^{\mathsf{3P}}_{\tau_1}\ge 1/2\}]\\ &\le \mathbb{P}^{(0,0)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{b^{\mathsf{3P}}_{\tau_1}\ge 1/2\} + \mathbb{P}^{(0,0)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{\tilde{b}^{\mathsf{3P}}_{\tau_1}\ge 1/2\}\\ &\le 2\mathbb{E}^{(0,0)}_{\mathbf{\Pi}^{\mathsf{3P}}}[b^{\mathsf{3P}}_{\tau_1}] + 2\mathbb{E}^{(0,0)}_{\mathbf{\Pi}^{\mathsf{3P}}}[\tilde{b}^{\mathsf{3P}}_{\tau_1}]\\ &\le 2[\mu\exp(-\xi\tau_1) + \tilde{\mu}\exp(-\tilde{\xi}\tau_1)]\\ &\le 4\max\{\mu,\tilde{\mu}\}\exp(-\min\{\xi,\tilde{\xi}\}\tau_1),\end{aligned} \tag{C.9}$$

where $b^{\mathsf{3P}}_{\tau_1}$ and $\tilde{b}^{\mathsf{3P}}_{\tau_1}$ denote the seller's beliefs under $\mathbf{\Pi}^{\mathsf{3P}}$ at the end of period τ_1, the first inequality follows by removing the terms with negative signs, the second inequality follows from Markov's inequality, the third inequality follows from Theorem 1 because $\mathbf{\Pi}^{\mathsf{3P}}$ employs a $(\boldsymbol{\delta}, t)$-discriminative policy for the first τ_1 periods, and the last inequality follows by rearranging terms. Now, regarding the first term on the right-hand side of (C.8), note that

$$\begin{aligned} J^{(j,k)}(\tau_1+\tau_2+1, T \,|\, \hat{\mathbf{\Pi}}) - \hat{V}^{(j,k)}_{\tau_1+\tau_2+1}\big((g_{\max}, y^{\mathsf{3P}}_{\tau_1+\tau_2+1})\big) \\ = \mathbb{E}^{(j,k)}_{\hat{\mathbf{\Pi}}}\big[\hat{V}^{(j,k)}_{\tau_1+\tau_2+1}\big((\hat{g}_{\tau_1+\tau_2+1}, \hat{y}_{\tau_1+\tau_2+1})\big)\big] - \hat{V}^{(j,k)}_{\tau_1+\tau_2+1}\big((g_{\max}, y^{\mathsf{3P}}_{\tau_1+\tau_2+1})\big) \\ \le \mathbb{E}^{(j,k)}_{\hat{\mathbf{\Pi}}}\big[\hat{V}^{(j,k)}_{\tau_1+\tau_2+1}\big((g_{\max}, \hat{y}_{\tau_1+\tau_2+1})\big)\big] - \hat{V}^{(j,k)}_{\tau_1+\tau_2+1}\big((g_{\max}, y^{\mathsf{3P}}_{\tau_1+\tau_2+1})\big) \le 0, \end{aligned} \tag{C.10}$$

where $\hat{y}_{\tau_1+\tau_2+1}$ and $\hat{g}_{\tau_1+\tau_2+1}$ are the inventory and goodwill levels, respectively, under $\hat{\mathbf{\Pi}}$ at the end of period $\tau_1+\tau_2+1$; the first inequality follows from Proposition 1(i) and the fact that $\hat{g}_{\tau_1+\tau_2+1} \le g_{\max}$; and the second inequality follows from Proposition 1(ii)-(iii) and the fact that $y^{\mathsf{3P}}_{\tau_1+\tau_2+1} \le 1$, which implies $\hat{V}^{(j,k)}_{\tau_1+\tau_2+1}\big((g_{\max}, y)\big) \le \hat{V}^{(j,k)}_{\tau_1+\tau_2+1}\big((g_{\max}, y^{\mathsf{3P}}_{\tau_1+\tau_2+1})\big)$ for all y, and therefore, $\mathbb{E}^{(j,k)}_{\hat{\mathbf{\Pi}}}[\hat{V}^{(j,k)}_{\tau_1+\tau_2+1}((g_{\max}, \hat{y}_{\tau_1+\tau_2+1}))] \le \hat{V}^{(j,k)}_{\tau_1+\tau_2+1}\big((g_{\max}, y^{\mathsf{3P}}_{\tau_1+\tau_2+1})\big)$.

Combining (C.8)-(C.10) and noting that $\mathbb{P}^{(j,k)}_{\mathbf{\Pi}^{\mathsf{3P}}}\{\mathcal{E}^{(j,k)}\} \le 1$, we deduce the following:

$$J^{(j,k)}(\tau_1+\tau_2+1, T \,|\, \hat{\mathbf{\Pi}}) - J^{(j,k)}(\tau_1+\tau_2+1, T \,|\, \mathbf{\Pi}^{\mathsf{3P}}) \le 4\Psi_1 \max\{\mu, \tilde{\mu}\} \exp\big(-\min\{\xi, \tilde{\xi}\}\tau_1\big) T = \Psi_2 \Psi_3^{-\tau_1} T,$$

where $\Psi_2 = 4\Psi_1 \max\{\mu, \tilde{\mu}\}$ and $\Psi_3 = \exp(\min\{\xi, \tilde{\xi}\})$. Q.E.D.

Appendix D: Analysis for the Extension to General Number of Hypotheses

In this section, we present the technical details and proofs for the generalized setting of Section 6. We first extend the belief update equations (4) and (6) to accommodate a general number of hypotheses. We start with the belief on the advertising model. At the beginning of period t, the seller updates her belief regarding $\lambda(\cdot)$ as follows:

$$\tilde{b}^{(k)}_t = \frac{\tilde{b}^{(k)}_{t-1} l_{\bar{N}}(g_{k,t}, N_t)}{\sum_{k' \in \mathcal{K}} b^{(k')}_{t-1} l_{\bar{N}}(g_{k',t}, N_t)}, \tag{D.1}$$

where $l_M(x, m) = x^m (1-x)^{M-m}$ for $M \in \mathbb{N}$, $x \in [0,1]$, and $m \in \{0, \ldots, M\}$. As in Section 2, we introduce $b^{(j)}_{t,c}$ to represent the seller's belief regarding $\rho(\cdot)$ before the purchase decision of the c-th customer in period t. Using this notation, the transition of $\{b^{(j)}_t\}$ is as follows:

$$b^{(j)}_{t,1} = b^{(j)}_{t-1}, \tag{D.2}$$

$$b^{(j)}_{t,c+1} = \mathbb{1}_{\{z_{t,c}=0\}} b^{(j)}_{t,c} + \mathbb{1}_{\{z_{t,c}\ge 1\}} \frac{b^{(j)}_{t,c} l_1(\rho_j(p_t), S_{t,c})}{\sum_{j' \in \mathcal{J}} b^{(j')}_{t,c} l_1(\rho_{j'}(p_t), S_{t,c})} \quad \text{for } c = 1, \ldots, N_t, \tag{D.3}$$

$$b^{(j)}_t = b^{(j)}_{t, N_t+1}. \tag{D.4}$$

We next provide a counterpart Lemma 1 for the extended setting of Section 6.

LEMMA 7. *Fix $\delta_g, \delta_C \in (0,1)$. Assume without loss of generality that $\lambda_k(a) > \lambda_{k-1}(a)$ for $k = 1, \ldots, K$ and $a \ge a_{\text{int}}$, where $a_{\text{int}} \in [a_{\min}, a_{\max}]$. With an inventory policy ensuring positive inventory* *for all periods, any sequence of advertising expenses* $\{a_\ell, \ell = 1, \ldots, t\}$ *with* $a_\ell \geq a_{\text{int}}$ *satisfying the following conditions also satisfy* (22) *and* (24)*:*

$$\lambda_k(a_\ell) - \lambda_{k-1}(a_\ell) \geq \delta_g \quad \text{for } k = 1, \ldots, K \text{ and } \ell = 1, \ldots, t, \tag{D.5}$$

$$\lambda_k(a_\ell) \geq \max\left\{(1-\beta)[1-(1-\delta_C)^{\frac{1}{N}}], 1-(1-\delta_C)^{\frac{1}{N}} - \beta g_0\right\} \quad \text{for } k \in \mathcal{K} \text{ and } \ell = 1, \ldots, t, \tag{D.6}$$

$$\lambda_k(a_\ell) \leq (1-\beta) g_{\max} \quad \text{for } k \in \mathcal{K} \text{ and } \ell = 1, \ldots, t. \tag{D.7}$$

The proof of this lemma is deferred to the end of this section. Before proceeding with our policy, we introduce some new notation. For all $i \in \{0, \ldots, J\}$, let e_i be the $(i+1)$-st standard basis vector in $\mathbb{R}^{J+1}$, and for all $i \in \{0, \ldots, K\}$, let $\tilde{e}_i$ be the $(i+1)$-st standard basis vector in $\mathbb{R}^{K+1}$. Furthermore, we denote by $\|\boldsymbol{x}\|$ the infinity norm of any vector $\boldsymbol{x}$, and for $\varepsilon \in [0, 1/2)$, we define $B_i(\varepsilon) \triangleq \{\boldsymbol{x} \in \mathcal{B}_J : \|\boldsymbol{x} - e_i\| \leq \varepsilon\}$ and $\tilde{B}_i(\varepsilon) \triangleq \{\boldsymbol{x} \in \mathcal{B}_K : \|\boldsymbol{x} - \tilde{e}_i\| \leq \varepsilon\}$.

Joint learning policy $\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}$ for general number of hypotheses

Initialization.

Let $\varepsilon \in [0, 1/2)$, $\tau \in \{1, \ldots, T\}$, and $\boldsymbol{\delta} = (\delta_g, \delta_\rho, \delta_C)$ such that $\delta_g, \delta_\rho, \delta_C \in (0, 1)$ and (18) is satisfied.

Joint experimentation with advertising, pricing, and inventory.

Assuming without loss of generality that $\lambda_k(a) > \lambda_{k-1}(a)$ for $k \in \{1, \ldots, K\}$ and $a \geq a_{\text{int}}$, where $a_{\text{int}} \in [a_{\min}, a_{\max}]$, set $\{a_\ell^{\mathtt{JL}}, \ell = 1, \ldots, \tau\}$ satisfying $a_\ell^{\mathtt{JL}} \in [a_{\text{int}}, a_{\max}]$ for $\ell = 1, \ldots, \tau$ and (D.5)-(D.7) for $t = \tau$.

Set $\{p_\ell^{\mathtt{JL}}, \ell = 1, \ldots, \tau\}$ satisfying (23) for $\ell = 1, \ldots, \tau$.

Choose $a_\ell = a_\ell^{\mathtt{JL}}$, $p_\ell = p_\ell^{\mathtt{JL}}$, and $q_\ell = \mathbb{1}_{\{y_{\ell-1}=0\}}$ for $\ell = 1, \ldots, \tau$.

Bayesian inference.

Compute $b_\tau^{(j)}$ and $\tilde{b}_\tau^{(k)}$ via (D.1)-(D.4), and set $\hat{\jmath} = \sum_{j \in \mathcal{J}} j \mathbb{1}_{\{\boldsymbol{b}_\tau \in B_j(\varepsilon)\}}$ and $\hat{k} = \sum_{k \in \mathcal{K}} k \mathbb{1}_{\{\tilde{\boldsymbol{b}}_\tau \in \tilde{B}_k(\varepsilon)\}}$.

Joint optimization of advertising, pricing, and inventory.

Compute the optimal policy $\hat{\boldsymbol{\Pi}} = (\hat{\boldsymbol{\pi}}^a, \hat{\boldsymbol{\pi}}^p, \hat{\boldsymbol{\pi}}^q)$ for the dynamic program (10) with $t = \tau + 1$, $j = \hat{\jmath}$, and $k = \hat{k}$. Choose $\pi_\ell^{a,\mathtt{JL}} = \hat{\pi}_\ell^a$, $\pi_\ell^{p,\mathtt{JL}} = \hat{\pi}_\ell^p$, and $\pi_\ell^{q,\mathtt{JL}} = \hat{\pi}_\ell^q$ for $\ell = \tau + 1, \ldots, T$.

Analogous to the joint learning policy $\boldsymbol{\Pi}^{\mathtt{JL}}$ in the binary prior case, the generalized policy $\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}$ employs a $(\boldsymbol{\delta}, t)$-discriminative policy—adapted to the extended conditions in Section 6—during the experimentation phase. The hypothesis $(\hat{\jmath}, \hat{k})$ for the subsequent joint optimization phase is identified as the one on which the seller's posterior belief is nearly concentrated, specifically when $\boldsymbol{b}_\tau \in B_{\hat{\jmath}}(\varepsilon)$ and $\tilde{\boldsymbol{b}}_\tau \in \tilde{B}_{\hat{k}}(\varepsilon)$. If one or both conditions are not met, $\hat{\jmath}$ and/or $\hat{k}$ are set to 0. With this adjustment, the policy retains the same logarithmic regret order in T with an experimentation phase of length τ, as stated in Theorem 3, whose proof follows.

Proof of Theorem 3. The proof follows two main steps. In the first step, we establish that $\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}$ attains exponentially fast learning of the individual demand and advertising response functions (analogous to Theorem 1). In the second step, we derive an upper bound on regret (analogous to Theorem 2).

Step 1(a). We start with the belief on the advertising response function. Iterating the belief update equation (D.1) for t times, it follows that, for each $k \in \mathcal{K}$,

$$\tilde{b}_t^{(k)} = \left[1 + \sum_{k' \in \mathcal{K}\setminus\{k\}} \frac{\tilde{b}_0^{(k')}}{\tilde{b}_0^{(k)}} \exp\left(\sum_{\ell=1}^{t} \log \frac{l_{\bar{N}}(g_{k',\ell}, N_\ell)}{l_{\bar{N}}(g_{k,\ell}, N_\ell)}\right)\right]^{-1}.$$

Fix $j \in \mathcal{J}$, $k \in \mathcal{K}$, and $k' \in \mathcal{K} \setminus \{k\}$. We deduce from (22) and Lemma A.2 of Harrison et al. (2012) that $\sum_{\ell=1}^{t} \log \frac{l_{\bar{N}}(g_{k',\ell},N_\ell)}{l_{\bar{N}}(g_{k,\ell},N_\ell)} \geq \sum_{\ell=1}^{t} \big((N_\ell - \bar{N} g_{k',\ell}) \log \frac{g_{k',\ell}(1-g_{k,\ell})}{g_{k,\ell}(1-g_{k',\ell})} + 2\bar{N}\delta_g^2\big)$. Consequently, letting $\tilde{\zeta}_t^{(k)} \triangleq \sum_{\ell=1}^{t} (N_\ell - \bar{N} g_{k',\ell}) \log \frac{g_{k',\ell}(1-g_{k,\ell})}{g_{k,\ell}(1-g_{k',\ell})}$, we obtain the following:

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k')}[\tilde{b}_t^{(k)}] &\leq \frac{1}{1 + \sum_{k' \in \mathcal{K}\setminus\{k\}} \frac{\tilde{b}_0^{(k')}}{\tilde{b}_0^{(k)}} \exp(\bar{N}\delta_g^2 t)} + \mathbb{P}_{\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k')}\{|\tilde{\zeta}_t^{(k)}| \geq \bar{N}\delta_g^2 t\} \\
&\leq \exp(-\bar{N}\delta_g^2 t) \frac{\tilde{b}_0^{(k)}}{1 - \tilde{b}_0^{(k)}} + \mathbb{P}_{\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k')}\{|\tilde{\zeta}_t^{(k)}| \geq \bar{N}\delta_g^2 t\}.
\end{aligned}$$

Based on the arguments used to derive (B.6) in the proof of Theorem 1, we deduce that $\tilde{\zeta}_t^{(k)}$ is a martingale under hypothesis (j,k'), as each N_ℓ has a binomial distribution with mean $\bar{N} g_{k',\ell}$. Thus, using Azuma's inequality (Ross 1996, p. 307), we derive an upper bound on $\mathbb{P}_{\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k')}\{|\tilde{\zeta}_t^{(k)}| \geq \bar{N}\delta_g^2 t\}$ and hence deduce that $\mathbb{E}_{\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k')}[\tilde{b}_t^{(k)}] \leq \tilde{\mu}_k \exp(-\tilde{\xi} t)$, where $\tilde{\mu}_k \triangleq \max\left\{2, \frac{\tilde{b}_0^{(k)}}{1-\tilde{b}_0^{(k)}}\right\}$ and $\tilde{\xi}$ is as in (B.7).

Step 1(b). Now, we consider the belief on the individual demand function and derive an upper bound on $\mathbb{E}_{\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j',k)}[b_t^{(j)} \,|\, \mathcal{A}_{t,m}^L]$, where $\mathcal{A}_{t,m}^L$ is as in the proof of Theorem 1. Fix $j \in \mathcal{J}$, $j' \in \mathcal{J} \setminus \{j\}$, and $k \in \mathcal{K}$. Using the belief update equations (D.2)-(D.4) repeatedly, we obtain the following:

$$\mathbb{E}_{\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j',k)}[b_t^{(j)} \,|\, \mathcal{A}_{t,m}^L] = \mathbb{E}_{\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j',k)}\left[\left(1 + \sum_{j' \in \mathcal{J}\setminus\{j\}} \frac{b_0^{(j')}}{b_0^{(j)}} \exp\left(\sum_{\ell \in \mathcal{L}_t} \sum_{c=1}^{C_\ell} \log \frac{l_1(\rho_{j'}(p_\ell), S_{\ell,c})}{l_1(\rho_j(p_\ell), S_{\ell,c})}\right)\right)^{-1} \,\middle|\, \mathcal{A}_{t,m}^L\right].$$

For the argument of the preceding exponential function, we deduce from (23) and Lemma A.2 of Harrison et al. (2012) that

$$\sum_{\ell \in \mathcal{L}_t} \sum_{c=1}^{C_\ell} \log \frac{l_1(\rho_{j'}(p_\ell), S_{\ell,c})}{l_1(\rho_j(p_\ell), S_{\ell,c})} \geq \sum_{\ell \in \mathcal{L}_t} \left[2\delta_\rho^2 + \sum_{c=1}^{C_\ell} [S_{\ell,c} - \rho_{j'}(p_\ell)] \log \frac{\rho_{j'}(p_\ell)(1-\rho_j(p_\ell))}{\rho_j(p_\ell)(1-\rho_{j'}(p_\ell))}\right].$$

Recall that on $\mathcal{A}_{t,m}^L$, $\sum_{\ell=1}^{t} \mathbb{1}_{\{C_\ell \geq 1\}} = m$, which yields the following upper bound:

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j',k)}[b_t^{(j)} \,|\, \mathcal{A}_{t,m}^L] \\
&\leq \mathbb{E}_{\boldsymbol{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j',k)}\left[\left(1 + \sum_{j' \in \mathcal{J}\setminus\{j\}} \frac{b_0^{(j')}}{b_0^{(j)}} \exp\left(2\delta_\rho^2 m + \sum_{\ell \in \mathcal{L}_t} \sum_{c=1}^{C_\ell} [S_{\ell,c} - \rho_{j'}(p_\ell)] \log \frac{\rho_{j'}(p_\ell)(1-\rho_j(p_\ell))}{\rho_j(p_\ell)(1-\rho_{j'}(p_\ell))}\right)\right)^{-1} \,\middle|\, \mathcal{A}_{t,m}^L\right]
\end{aligned}$$

$$\leq \frac{1}{1+\frac{1-b_0^{(j)}}{b_0^{(j)}}\exp\left(\delta_\rho^2 m\right)} + \mathbb{P}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j',k)}\left\{\left|\sum_{\ell\in\mathcal{L}_t}\sum_{c=1}^{C_\ell}[S_{\ell,c}-\rho_{j'}(p_\ell)]\log\frac{\rho_{j'}(p_\ell)(1-\rho_j(p_\ell))}{\rho_j(p_\ell)(1-\rho_{j'}(p_\ell))}\right| \geq \delta_\rho^2 m \,\middle|\, \mathcal{A}_{t,m}^L\right\},$$

where the second inequality follows by considering the cases where the sum inside the preceding exponential function, $\sum_{\ell\in\mathcal{L}_t}\sum_{c=1}^{C_\ell}[S_{\ell,c}-\rho_{j'}(p_\ell)]\log\frac{\rho_{j'}(p_\ell)(1-\rho_j(p_\ell))}{\rho_j(p_\ell)(1-\rho_{j'}(p_\ell))}$, has an absolute value smaller or larger than $\delta_\rho^2 m$. Moreover, to find an upper bound on the right-hand side of the second inequality, we repeat the arguments used in Step 2 of in the proof of Lemma 5, and deduce that

$$\mathbb{E}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j',k)}[b_t^{(j)}\,|\,\mathcal{A}_{t,m}^L] \leq \frac{1}{1+\frac{1-b_0^{(j)}}{b_0^{(j)}}\exp\left(\delta_\rho^2 m\right)} + 2\exp\left(-\frac{\delta_\rho^4 m}{2\bar{N}^2\gamma}\right) \leq \mu_j \exp\left(-\frac{\delta_\rho^4 m}{2\bar{N}^2\gamma}\right),$$

where $\mu_j \triangleq \max\left\{2, \frac{b_0^{(j)}}{1-b_0^{(j)}}\right\}$. Finally, because $\mathbb{P}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j',k)}\{\mathcal{A}_{t,m}^L\} \leq (1-\delta_C)^{t-m}$, we conclude that $\mathbb{E}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j',k)}[b_t^{(j)}] \leq \sum_{m=0}^{t}\binom{t}{m}(1-\delta_C)^{t-\ell}\mu_j\exp\left(-\frac{\delta_\rho^4 m}{2\bar{N}^2\gamma}\right) = \mu_j\exp\left(-\xi t\right)$, where ξ is as in (B.2).

Step 2. We now derive an upper bound on the regret $\mathsf{R}^{(j,k)}\left(\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}\right)$. This derivation parallels that of Theorem 2, as extending to a general number of hypotheses requires no substantive modifications except in part (iii) of Lemma 6, which must be adjusted to reflect the enlarged hypothesis space and the revised selection rule for $(\hat{\jmath},\hat{k})$. To avoid redundancy, we present only the updated proof for this part.

Define $\mathcal{E}_m^{(j,k)} = \{1-b_{\tau_1}^{(j)} \leq \varepsilon\} \cap \{1-\tilde{b}_{\tau_1}^{(k)} \leq \varepsilon\}$. On $\mathcal{E}_m^{(j,k)}$, hypothesis (j,k) is selected because $\boldsymbol{b}_{\tau_1} \in B_j(\varepsilon)$ when $1-b_{\tau_1}^{(j)} \leq \varepsilon$ and $\tilde{\boldsymbol{b}}_{\tau_1} \in \tilde{B}_k(\varepsilon)$ when $1-\tilde{b}_{\tau_1}^{(k)} \leq \varepsilon$. In what follows, we bound $1-\mathbb{P}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k)}\{\mathcal{E}_m^{(j,k)}\}$, which upper bounds the probability that the true hypothesis (j,k) is not selected:

$$\begin{aligned}
1-\mathbb{P}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k)}\{\mathcal{E}_m^{(j,k)}\} &= 1-[1-\mathbb{P}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k)}\{1-b_{\tau_1}^{(j)}>\varepsilon\}][1-\mathbb{P}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k)}\{1-\tilde{b}_{\tau_1}^{(k)}>\varepsilon\}] \\
&\overset{(a)}{\leq} \mathbb{P}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k)}\{1-b_{\tau_1}^{(j)}>\varepsilon\} + \mathbb{P}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k)}\{1-\tilde{b}_{\tau_1}^{(k)}>\varepsilon\} \\
&\overset{(b)}{\leq} \frac{\mathbb{E}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k)}[1-b_{\tau_1}^{(j)}]}{\varepsilon} + \frac{\mathbb{E}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k)}[1-\tilde{b}_{\tau_1}^{(k)}]}{\varepsilon} \\
&\overset{(c)}{=} \frac{\sum_{j'\in\mathcal{J}\setminus\{j\}}\mathbb{E}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k)}[b_{\tau_1}^{(j')}]}{\varepsilon} + \frac{\sum_{k'\in\mathcal{K}\setminus\{k\}}\mathbb{E}_{\mathbf{\Pi}_{\mathtt{G}}^{\mathtt{JL}}}^{(j,k)}[\tilde{b}_{\tau_1}^{(k')}]}{\varepsilon} \\
&\overset{(d)}{\leq} \frac{\sum_{j'\in\mathcal{J}\setminus\{j\}}\mu_{j'}\exp(-\xi\tau_1)}{\varepsilon} + \frac{\sum_{k'\in\mathcal{K}\setminus\{k\}}\tilde{\mu}_{k'}\exp(-\tilde{\xi}\tau_1)}{\varepsilon} \\
&\overset{(e)}{\leq} \frac{2\max\{(J+1)\max_{j'\in\mathcal{J}}\mu_{j'},(K+1)\max_{k'\in\mathcal{K}}\mu_{k'}\}}{\varepsilon}\exp(-\min\{\xi,\tilde{\xi}\}\tau_1),
\end{aligned}$$

where (a) follows by removing the terms with negative signs, (b) follows from Markov's inequality, (c) follows because $\sum_{j'\in\mathcal{J}} b_{\tau_1}^{(j')} = 1$ and $\sum_{k'\in\mathcal{K}}\tilde{b}_{\tau_1}^{(k')} = 1$, (d) follows from the arguments used in Step 1 of the current proof, and (e) follows by rearranging terms. As a result, Lemma 6 continues to hold in the generalized setting of Section 6, except that Ψ_2 in part (iii) is replaced by

$$\Psi_2(\mathcal{J},\mathcal{K}) = \frac{2\max\{(J+1)\max_{j'\in\mathcal{J}}\mu_{j'},(K+1)\max_{k'\in\mathcal{K}}\mu_{k'}\}}{\varepsilon}.$$

Accordingly, $\mu^{\mathtt{JL}}$, which depends on Ψ_2, is replaced by $\mu^{\mathtt{JL}}_{\mathtt{G}}$, defined as

$$\mu^{\mathtt{JL}}_{\mathtt{G}} \triangleq 3\max\left\{\frac{\Psi_1}{\log\Psi_3}, \Psi_1, 2\Psi_2(\mathcal{J},\mathcal{K})\right\}. \tag{D.8}$$

This completes the proof and establishes that the logarithmic regret bound continues to hold under an arbitrary number of hypotheses. Q.E.D.

Proof of Lemma 7. We prove this result by induction. For the base step, consider $\ell = 1$. By (1), $g_{k,1} = \beta g_0 + \lambda_k(a_1)$ for $k \in \mathcal{K}$, and by (D.5), $\lambda_k(a_1) - \lambda_{k-1}(a_1) \geq \delta_g$ for $k = 1,\ldots,K$. In addition, (D.7) implies that $g_{\max} \geq \beta g_0 + \lambda_k(a_1)$ for $k \in \mathcal{K}$. Consequently,

$$g_{k,1} - g_{k-1,1} = \lambda_k(a_1) - \lambda_{k-1}(a_1) \geq \delta_g \quad \text{for } k = 1,\ldots,K.$$

Hence, $g_{k,1}$ is monotone in $k \in \mathcal{K}$ with an increment higher than δ_g so that (22) holds for $\ell = 1$. Because there is positive inventory in period 1, $\mathbb{P}^{(j,k)}\{C_1 \geq 1\} = 1 - (1 - g_{k,1})^{\bar{N}}$. Therefore, when $g_{k,1} \geq 1 - (1-\delta_C)^{1/\bar{N}}$, (24) is satisfied for $\ell = 1$. Moreover, since a_1 satisfies (D.6), we have that $g_{k,1} \geq 1 - (1-\delta_C)^{1/\bar{N}}$, and thus, (24) holds for $\ell = 1$.

For the induction step, consider period $\ell > 1$, and suppose that $g_{\max} \geq g_{k,\ell-1}$, $g_{k,\ell-1} - g_{k-1,\ell-1} \geq \delta_g$ for $k = 1,\ldots,K$, and $g_{k,\ell-1} \geq 1 - (1-\delta_C)^{1/\bar{N}}$ for $k \in \mathcal{K}$. Based on (1), (D.7) and the fact that $g_{\max} \geq g_{k,\ell-1}$, it follows that $g_{k,\ell} = \beta g_{k,\ell-1} + \lambda_k(a_\ell)$. Consequently, we verify (22) as follows:

$$|g_{k,\ell} - g_{k-1,\ell}| = \beta(g_{k,\ell-1} - g_{k-1,\ell-1}) + \lambda_k(a_\ell) - \lambda_{k-1}(a_\ell) \geq \beta\delta_g + (1-\beta)\delta_g = \delta_g.$$

As a result, since there is positive inventory in each period, the induction hypothesis that $g_{k,\ell-1} \geq 1 - (1-\delta_C)^{1/\bar{N}}$ and (D.6) jointly imply that (24) is satisfied. Q.E.D.

Appendix E: Further Details on Numerical Results

The numerical analysis in Section 7 is based on the parameter values reported in Table 1.

Table 1 Parameter Values for Numerical Results in Section 7

$p_{\min}$	$p_{\max}$	h	$\bar{N}$	$a_{\min}$	$a_{\max}$	c_a	β	$q_{\min}$	$q_{\max}$	g_0	$g_{\max}$
0.5	1.5	0.25	3	0	4	0.1	0.85	0	3	0.99	0.99

Table 2 displays the regret of the joint learning policy under hypothesis $(1,1)$ for different values of τ and T in this setting. Table 3 displays the passive learning policy's regret in the same setting. Comparing Tables 2 and 3 emphasizes the advantages of using the joint learning policy as illustrated in Figure 4. In particular, the joint learning policy's regret is much smaller than that of the passive learning policy.

Table 2 $\mathsf{R}^{(1,1)}\left(\boldsymbol{\Pi}^{\mathrm{JL}}\right)$: Regret of Joint Learning Policy under Hypothesis $(1,1)$

	$\tau=3$	$\tau=5$	$\tau=10$
$T=200$	6.31	5.34	7.23
$T=400$	10.12	7.23	8.10
$T=600$	14.23	10.00	8.85
$T=800$	16.57	12.20	9.45
$T=1000$	23.22	14.17	10.81
$T=1200$	29.84	15.49	12.04
$T=1400$	30.29	17.14	13.29
$T=1600$	36.15	21.12	14.27
$T=1800$	41.76	23.48	15.92
$T=2000$	45.95	25.67	16.43

Table 3 $\mathsf{R}^{(1,1)}\left(\boldsymbol{\Pi}^{\mathrm{PL}}\right)$: Regret of Passive Learning Policy under Hypothesis $(1,1)$

	$\tau=5$	$\tau=10$
$T=200$	62.40	50.79
$T=400$	126.18	101.57
$T=600$	189.60	152.34
$T=800$	252.75	203.49
$T=1000$	313.00	253.07
$T=1200$	377.10	304.78
$T=1400$	435.57	357.22
$T=1600$	507.32	408.37
$T=1800$	568.20	453.10
$T=2000$	641.74	506.52

Table 2 reports regret when the chosen price and advertising expense satisfy the $(\boldsymbol{\delta},t)$-discriminative conditions in (12)-(14). To assess the consequences of violating these conditions, Table 4 examines the joint learning policy's regret performance when the parameters are non-discriminative.

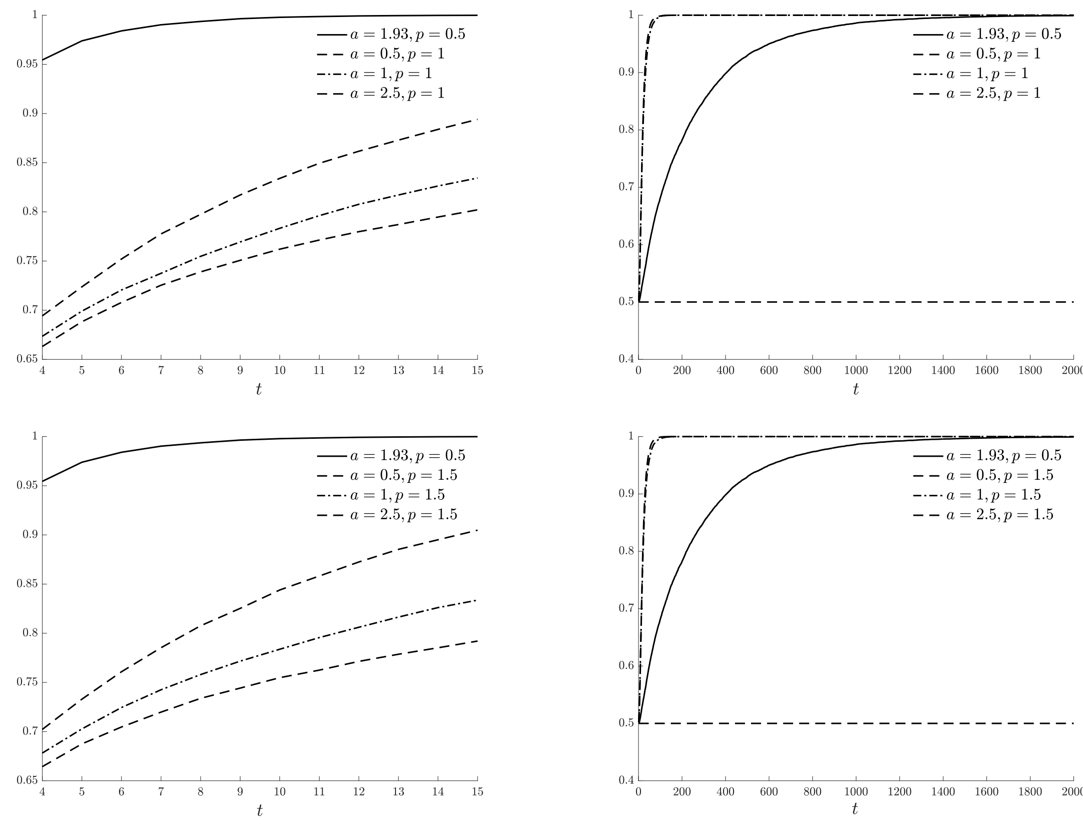
Table 4 $\mathsf{R}^{(1,1)}\left(\boldsymbol{\Pi}^{\mathrm{JL}}\right)$: Regret of Joint Learning Policy under Hypothesis $(1,1)$ for Non-discriminative Parameters

	$(p,a)=(1,0.5)$	$(p,a)=(1.5,0.5)$	$(p,a)=(1,1)$	$(p,a)=(1.5,1)$	$(p,a)=(1,2.5)$	$(p,a)=(1.5,2.5)$
$T=200$	34.93	39.85	28.94	37.73	24.36	31.06
$T=400$	66.06	72.26	57.48	64.93	47.12	54.92
$T=600$	97.08	101.33	89.21	91.60	69.07	75.52
$T=800$	127.45	132.76	109.91	120.94	87.95	100.94
$T=1000$	171.88	172.48	141.57	150.77	113.34	123.65
$T=1200$	189.88	193.00	168.43	180.52	141.28	152.85
$T=1400$	213.38	238.90	196.23	199.14	160.16	166.19
$T=1600$	259.10	264.36	218.04	229.98	183.29	202.96
$T=1800$	285.72	289.32	254.14	264.65	208.24	209.19
$T=2000$	309.51	325.57	283.97	286.05	223.13	237.83

In Table 4, we consider different price and advertising expense values that the joint learning policy uses during the first 10 periods. These regret performances are also illustrated in Figure 3. As shown in that figure and in Table 4, the joint learning policy performs poorly when the price and advertising expense do not satisfy the $(\boldsymbol{\delta},t)$-discriminative conditions in (12)-(14).

To explain this poor performance, we examine the learning dynamics under the different price-advertising pairs. The $(\boldsymbol{\delta},t)$-discriminative conditions are designed to guarantee exponentially fast learning for both the advertising response and demand functions, and their violation undermines this property. In Figure 5, we illustrate the expected beliefs for the price and advertising expense pairs that generate the regret performances in Table 4, and the pair that satisfies $(\boldsymbol{\delta}, t)$-discriminative conditions.

Figure 5 The effect of different price and advertising expenses on learning



(a) expected value of b_t when $\rho(\cdot) = \rho_1(\cdot)$ (b) expected value of $\tilde{b}_t$ when $\lambda(\cdot) = \lambda_1(\cdot)$

Figure 5 highlights why joint discriminative design is essential for achieving low regret. Lower advertising expenses (e.g., $a = 0.5$ or $a = 1$) may accelerate learning of the advertising response function, but they simultaneously reduce the number of customers exposed to price variation, slowing down demand learning. By contrast, a $(\boldsymbol{\delta}, t)$-discriminative pair achieves balanced, simultaneous learning of both functions, which translates into superior regret performance.

References

Harrison JM, Keskin NB, Zeevi A (2012) Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science* 58(3):570–586.

Levi R, Perakis G, Uichanco J (2015) The data-driven newsvendor problem: New bounds and insights. *Operations Research* 63(6):1294–1306.

Massart P (2007) *Concentration Inequalities and Model Selection: Lectures from the 33rd Summer School on Probability Theory in Saint-Flour, July 6–23, 2003* (Springer, New York).

Ross SM (1996) *Stochastic Processes*, volume 2 (Wiley, New York).