

E-Companion

EC.1. Proof of Theorem 1

THEOREM 1. *The optimal expected revenue function $G^*(Q_1, Q_2, T)$ is monotone in Q_1 , Q_2 and T . Hence,*

1. $G^*(Q_1 + 1, Q_2, T) \geq G^*(Q_1, Q_2, T)$,
2. $G^*(Q_1, Q_2, T + 1) \geq G^*(Q_1, Q_2, T)$.

Proof of Theorem 1 Nb. throughout the proof we only consider the cases where inventory for both products is positive for all relevant states, because considering all boundary states would take a lot of trivial case distinctions.

We prove 1 by induction in t . For $t = 1$

$$G^*(q_1 + 1, q_2, t) = \max_p \{ \delta_1(p)p_1 + \delta_2(p)p_2 + \delta_O(p)p \} = G^*(q_1, q_2, t).$$

Suppose the statement is true up to $t - 1$, then

$$\begin{aligned} G^*(q_1 + 1, q_2, t) &= \max_p \{ \delta_1(p)[p_1 + G^*(q_1, q_2, t - 1)] + \delta_2(p)[p_2 + G^*(q_1 + 1, q_2 - 1, t - 1)] \\ &\quad + \delta_O(p)[p + \max\{G^*(q_1, q_2, t - 1), G^*(q_1 + 1, q_2 - 1, t - 1)\}] + \delta_\emptyset(p)G^*(q_1 + 1, q_2, t - 1) \} \\ &\geq \max_p \{ \delta_1(p)[p_1 + G^*(q_1 - 1, q_2, t - 1)] + \delta_2(p)[p_2 + G^*(q_1, q_2 - 1, t - 1)] \\ &\quad + \delta_O(p)[p + \max\{G^*(q_1 - 1, q_2, t - 1), G^*(q_1, q_2 - 1, t - 1)\}] + \delta_\emptyset(p)G^*(q_1, q_2, t - 1) \} \\ &= G^*(q_1, q_2, t), \end{aligned}$$

what completes the proof of 1.

We prove 2 by induction in t . For $t = 0$

$$G^*(q_1, q_2, t) = 0 \leq \max_p \{ \delta_1(p)p_1 + \delta_2(p)p_2 + \delta_O(p)p \} = G^*(q_1, q_2, t + 1).$$

Suppose the statement is true up to t , then

$$\begin{aligned} G^*(q_1, q_2, t + 1) &= \max_p \{ \delta_1(p)[p_1 + G^*(q_1 - 1, q_2, t)] + \delta_2(p)[p_2 + G^*(q_1, q_2 - 1, t)] \\ &\quad + \delta_O(p)[p + \max\{G^*(q_1 - 1, q_2, t - 1), G^*(q_1, q_2 - 1, t)\}] + \delta_\emptyset(p)G^*(q_1, q_2, t) \} \\ &\geq \max_p \{ \delta_1(p)[p_1 + G^*(q_1 - 1, q_2, t - 1)] + \delta_2(p)[p_2 + G^*(q_1, q_2 - 1, t - 1)] \\ &\quad + \delta_O(p)[p + \max\{G^*(q_1 - 1, q_2, t - 1), G^*(q_1, q_2 - 1, t - 1)\}] + \delta_\emptyset(p)G^*(q_1, q_2, t - 1) \} \\ &= G^*(q_1, q_2, t), \end{aligned}$$

what completes the proof of 2. $\square$

EC.2. Proof of Theorem 2

THEOREM 2. *The structure of the optimal solution for the (semi-) abundant state(s) is as follows:*

1. *The optimal opaque price is equal in all abundant states and it does not matter whether the opaque demand is fulfilled with product 1 or product 2.*
2. *In a j -abundant state it is always optimal to fulfill opaque demand with product j and the optimal opaque price in a j -abundant state is the same for all states with equal inventory levels for product $i \neq j$ and remaining number of time units.*
3. *The optimal opaque price strictly increases in the inventory of product $i \neq j$ in a j -abundant state, given that it is optimal to offer the opaque product in the state.*
4. *The optimal opaque price strictly decreases in the number of remaining time units, in a j -abundant state, given that it is optimal to offer the opaque product in the state.*

Proof of Theorem 2 The proof is via induction and contains the following (in)equalities:

1. $p^*(q_1, q_2, t) = p^*(q'_1, q'_2, t')$ and $O^*(q_1, q_2, t) = \{1, 2\}$ for all $q_1, q_2 \geq t$ and $q'_1, q'_2 \geq t'$.
2. $p^*(q_1, q_2, t) = p^*(q_1, q'_2, t)$ and $O^*(q_1, q_2, t) = O^*(q_1, q'_2, t) = 2$ for all $q_2, q'_2 \geq t$
3. $p^*(q_1, q_2, t) < p^*(q'_1, q_2, t)$ if $p^*(q'_1, q_2, t) < \bar{p}$ for all $q_2 \geq t$ and $1 \leq q_1 < q'_1 \leq t$
4. $p^*(q_1, q_2, t) > p^*(q_1, q_2, t')$ if $p^*(q_1, q_2, t') < \bar{p}$ for all $q_2 \geq t'$ and $1 \leq q_1 \leq t < t'$.
5. $\Delta_2 G(q_1, q_2, t) = 0$ for all $q_2 > t$,
6. $\Delta_1 G(q_1, q_2, t) = \Delta_1 G(q_1, q_2 + 1, t)$ for all $q_2 > t$,
7. $\Delta_1 G(q_1, q_2, t) > \Delta_1 G(q_1 + 1, q_2, t)$ for $1 \leq q_1 < t$ and $q_2 > t$,
8. $\Delta_1 G(q_1, q_2, t - 1) < \Delta_1 G(q_1, q_2, t)$ for $1 \leq q_1 < t$ and $q_2 > t$.

When $t = 1$ and $q_1, q_2 \geq 1$, the $\Delta_1 G$ and $\Delta_2 G$ term in (1) and (2) vanish and $G^*(q_1, q_2, t) = \max_p \{ \lambda \pi_1 (1 - \phi_1(p)) p_1 + \lambda \pi_2 (1 - \phi_2(p)) p_2 + (\lambda_O(p) + \lambda \pi_1 \phi_1(p) + \lambda \pi_2 \phi_2(p)) p \}$, $p^*(q_1, q_2, t) = \arg\max_p G^*(q_1, q_2, t)$ and $O^*(q_1, q_2, t) = \{1, 2\}$. This forms the induction basis for 1. and 2. We will refer to the max and argmax of this equation as G^A and p^A .

Together with the fact that $G^*(0, q_2, 1) = G^*(0, q'_2, 1) = (\lambda \pi_1 \xi_{12} + \lambda \pi_2) p_2 < \lambda \pi_1 p_1 + \lambda \pi_2 p_2$, this also is a basis for 5, and 6.

By adding $\Delta_1 G^*(1, q_2, 1) > \lambda \pi_1 p_1 + \lambda \pi_2 p_2 - (\lambda \pi_1 \xi_{12} + \lambda \pi_2) p_2 > 0$, we create a basis for 7.

Now we define $p^- := p^*(1, q_2, 1)$. Based on the above we can now obtain an induction basis for VIII:

$$\begin{aligned}
 \Delta_1 G^*(1, q_2, 1) - \Delta_1 G^*(1, q_2, 2) &= G^*(1, q_2, 1) - G^*(1, q_2, 2) - (\lambda \pi_1 \xi_{12} + \lambda \pi_2) \Delta_2 G^*(0, q_2, 1) \\
 &< -G^*(1, q_2, 1) + \lambda \pi_1 (1 - \phi_1(p^-)) \Delta_1 G^*(1, q_2, 1) \\
 &< 0.
 \end{aligned}$$

Because 5. is true for $t = 1$, we can prove a basis for 3. and 4. by showing $p^*(1, q_2, 2) < p^A$ if $p^A < \bar{p}$.

$$\begin{aligned} p^*(1, q_2, 2) &= \operatorname{argmax}_p \{ \lambda \pi_1 (1 - \phi_1(p)) (p_1 - \Delta_1 G^*(1, q_2, 1)) + \lambda \pi_2 (1 - \phi_2(p)) p_2 + (\lambda \pi_1 \phi_1(p) + \lambda \pi_2 \phi_2(p) + \lambda_O(p)) p \} \\ &< \operatorname{argmax}_p \{ \lambda \pi_1 (1 - \phi_1(p)) p_1 + \lambda \pi_2 (1 - \phi_2(p)) p_2 + (\lambda \pi_1 \phi_1(p) + \lambda \pi_2 \phi_2(p) + \lambda_O(p)) p \} \\ &= p^A, \end{aligned}$$

because $\phi_1(p)$ strictly decreases in p for all $p \leq p^A < \bar{p}$.

Assuming all statements to be true up to $t - 1$:

This implies that $\Delta_2 G^*(q_1, q_2, t - 1) = \Delta_2 G^*(q_1, q_2 + 1, t - 1) = 0$, $\Delta_1 G^*(q_1, q_2, t - 1) = \Delta_1 G^*(q_1, q_2 + 1, t - 1)$ and $2 \subseteq O^*(q_1, q_2, t)$ for $q_2 \geq t$. So

$$\begin{aligned} G^*(q_1, q_2, t) &= G^*(q_1, q_2, t - 1) + \max_p \{ \lambda \pi_1 (1 - \phi_1(p)) (p_1 - \Delta_1 G^*(q_1, q_2, t - 1)) + \lambda \pi_2 (1 - \phi_2(p)) p_2 \\ &\quad + (\lambda_O(p) + \lambda \pi_1 \phi_1(p) + \lambda \pi_2 \phi_2(p)) p \} \\ &= G^*(q_1, q_2 + 1, t - 1) + \max_p \{ \lambda \pi_1 (1 - \phi_1(p)) (p_1 - \Delta_1 G^*(q_1, q_2 + 1, t - 1)) + \lambda \pi_2 (1 - \phi_2(p)) p_2 \\ &\quad + (\lambda_O(p) + \lambda \pi_1 \phi_1(p) + \lambda \pi_2 \phi_2(p)) p \} \\ &= G^*(q_1, q_2 + 1, t), \end{aligned}$$

what proves 2. and 5. This also implies 6. because

$$\begin{aligned} \Delta_1 G^*(q_1, q_2, t) - \Delta_1 G^*(q_1, q_2 + 1, t) &= G^*(q_1, q_2, t) - G^*(q_1 - 1, q_2, t) - G^*(q_1, q_2 + 1, t) + G^*(q_1 - 1, q_2 + 1, t) \\ &= \Delta_2 G^*(q_1 - 1, q_2 + 1, t) - \Delta_2 G^*(q_1, q_2 + 1, t) \\ &= 0. \end{aligned}$$

When we add that $q_1 \geq t$, we know that $\Delta_1 G^*(q_1, q_2, t) = \Delta_1 G^*(q_1 + 1, q_2, t - 1) = 0$ and $O^*(q_1, q_2, t) = \{1, 2\}$, $G^*(q_1, q_2, t) = G^A$ and $p^*(q_1, q_2, t) = p^A$, what proves 1.

3. is true because

$$\begin{aligned} p^*(q_1, q_2, t) &= \operatorname{argmax}_p \{ \lambda \pi_1 (1 - \phi_1(p)) (p_1 - \Delta_1 G^*(q_1, q_2, t - 1)) + \lambda \pi_2 (1 - \phi_2(p)) p_2 + (\lambda \pi_1 \phi_1(p) + \lambda \pi_2 \phi_2(p) + \lambda_O(p)) p \} \\ &> \operatorname{argmax}_p \{ \lambda \pi_1 (1 - \phi_1(p)) (p_1 - \Delta_1 G^*(q_1 + 1, q_2, t - 1)) + \lambda \pi_2 (1 - \phi_2(p)) p_2 + (\lambda \pi_1 \phi_1(p) + \lambda \pi_2 \phi_2(p) + \lambda_O(p)) p \}, \end{aligned}$$

because $(1 - \phi_1(p))(\Delta_1 G^*(q_1, q_2, t - 1) - \Delta_1 G^*(q_1 + 1, q_2, t - 1))$ strictly increases in p for all $p < \bar{p}$.

Now we define $p^+ := p^*(q_1 + 1, q_2, t)$ and $p^- := p^*(q_1 - 1, q_2, t)$ and with the simple argument

$$\begin{aligned} \Delta_1 G^*(q_1, q_2, t) - \Delta_1 G^*(q_1 + 1, q_2, t) &> [1 - \lambda \pi_1 (1 - \phi_1(p^+))] (\Delta_1 G^*(q_1, q_2, t - 1) - \Delta_1 G^*(q_1 + 1, q_2, t - 1)) \\ &\quad + \lambda \pi_1 (1 - \phi_1(p^-)) (\Delta_1 G^*(q_1 - 1, q_2, t - 1) - \Delta_1 G^*(q_1, q_2, t - 1)) \\ &> 0, \end{aligned}$$

we conclude that 7. is true.

Along similar lines we prove 8. First we define $p^- := p^*(q_1, q_2, t-1)$ and $p^+ := p^*(q_1-1, q_2, t) < p^-$, because of 3. and 4. But then

$$\begin{aligned}
& \Delta_1 G^*(q_1, q_2, t-1) - \Delta_1 G^*(q_1, q_2, t) \\
& < \Delta_1 G^*(q_1, q_2, t-2) - \Delta_1 G^*(q_1, q_2, t-1) - \lambda\pi_1(1 - \phi_1(p^-))(\Delta_1 G^*(q_1, q_2, t-2) - \Delta_1 G^*(q_1-1, q_2, t-2)) \\
& \quad + \lambda\pi_1(1 - \phi_1(p^+))(\Delta_1 G^*(q_1, q_2, t-1) - \Delta_1 G^*(q_1-1, q_2, t-1)) \\
& < (1 - \lambda\pi_1(1 - \phi_1(p^+)))(\Delta_1 G^*(q_1, q_2, t-2) - \Delta_1 G^*(q_1, q_2, t-1)) \\
& \quad + \lambda\pi_1(1 - \phi_1(p^-))(\Delta_1 G^*(q_1-1, q_2, t-2) - \Delta_1 G^*(q_1-1, q_2, t-1)) \\
& < 0.
\end{aligned}$$

Last, we prove 4., with use of 8.:

$$\begin{aligned}
& p^*(q_1, q_2, t) \\
& = \operatorname{argmax}_p \{ \lambda\pi_1(1 - \phi_1(p))(p_1 - \delta_1 G^*(q_1, q_2, t-1)) + \lambda\pi_2(1 - \phi_2(p))p_2 + (\lambda\pi_1\phi_1(p) + \lambda\pi_2\phi_2(p) + \lambda_O(p))p \} \\
& > \operatorname{argmax}_p \{ \lambda\pi_1(1 - \phi_1(p))(p_1 - \delta_1 G^*(q_1, q_2, t)) + \lambda\pi_2(1 - \phi_2(p))p_2 + (\lambda\pi_1\phi_1(p) + \lambda\pi_2\phi_2(p) + \lambda_O(p))p \} \\
& = p^*(q_1, q_2, t+1)
\end{aligned}$$

if $p^*(q_1, q_2, t+1) < \bar{p}$, again because $\phi_1(p)$ strictly decreases in p for all $p \leq \bar{p}$. $\square$

EC.3. Proof of Theorem 3

THEOREM 3. *If there are functions $g_1(\cdot)$ and $g_2(\cdot)$ such that policy φ satisfies the following properties, for each repetition k ,*

1. $\mathbb{E} \left[\|\beta - \hat{\beta}_k\|_\infty^2 \right] \leq g_1(k),$
2. $\sum_{l=1}^k \mathbb{E} \left[\|p^*(\hat{\beta}_l) - p_l\|_1 \right] \leq g_2(k),$
3. $O_k = O^*(\hat{\beta}_k),$

where $\hat{\beta}_k$ is an estimator for β based on $\mathcal{H}_{T_k}$, (p_k, O_k) is a vector of length T that follows from policy φ and contains the decisions applied within horizon k and $(p^*(\hat{\beta}_k), O^*(\hat{\beta}_k))$ are the perceived optimal decision vectors (of length T) within repetition k given the parameter estimates at the beginning of repetition, then $\mathbf{R}_k^\varphi = O(\max\{\sum_{l=1}^k g_1(l), g_2(k)\})$.

Proof of Theorem 3 The outline of the proof is as follows:

$$\begin{aligned}
\sum_{l=1}^k \mathbb{E} [G^* - G(p_l, O_l)] & \leq \sum_{l=1}^k \mathbb{E} \left[G^* - G(p^*(\hat{\beta}_l), O^*) + |G(p^*(\hat{\beta}_l), O^*) - G(p_l, O^*)| + |G(p_l, O^*) - G(p_l, O_l)| \right] \\
& \stackrel{A+B+C}{\leq} \sum_{l=1}^k \mathbb{E} \left[c_0 \|\beta - \hat{\beta}_l\|_\infty^2 + c_1 \|p^*(\hat{\beta}_l) - p_l\|_1 + c_2 \|O^* - O_l\|_1^2 \right]
\end{aligned}$$

$$\begin{aligned}
&\stackrel{2.+3.+D}{\leq} \sum_{l=1}^k \mathbb{E} \left[c_3 \|\beta - \hat{\beta}_l\|_\infty^2 \right] + c_1 g_2(k) \\
&\stackrel{1.}{\leq} c_3 \sum_{l=1}^k g_1(l) + c_1 g_2(k),
\end{aligned}$$

where G^* is short for $G^*(Q_1, Q_2, T)$, $G(p_k, O_k)$ is short for $G(Q_1, Q_2, T, p_k, O_k)$ what is used to denote the expected revenue as if decisions p_k and O_k were applied, and $\{c_i\}$ are constants. Inequalities A, B, C and D are shown one by one below, but we first need one auxiliary lemma.

LEMMA EC.1. *For a fixed opaque product decision vector O , there exists a constant $c > 0$ such that:*

$$\|p^*(q_1, q_2, t, \beta) - p^*(q_1, q_2, t, \beta')\|_1 \leq c \|\beta - \beta'\|_\infty, \quad (\text{EC.1})$$

where $p^*(q_1, q_2, t, \beta)$ and $p^*(q_1, q_2, t, \beta')$ are the optimal opaque prices in state (q_1, q_2, t) if the parameter matrix is β or β' respectively.

Proof We define the function

$$f_{\beta, C}(p) = (p_1 - c_1)h_1(\beta_0^1 + \beta_{1,1}p) + (p_2 - c_2)h_2(\beta_0^2 + \beta_{1,1}^2p) + (p - c_O)h_O(\beta_0^O + \beta_{1,1}^Op).$$

$f_{\beta, C}$ is twice continuous differentiable in all elements of β and $C = (c_1, c_2, C_O)$, because $h_d(\cdot)$ is a twice continuous differentiable function. The argmaximum of $f_{\beta, C}(p)$ is assumed to be unique, and is referred to as $p^*(\beta, C)$, and to $f_{\beta, C}(p^*(\beta, C))$ as $f(\beta, C)$. $p^*(\beta, C)$ and $f(\beta, C)$ are also twice continuous differentiable in all elements of β and C .

We refer to any particular element of β or C as d , and we refer to $f(\beta, C, d)$ as the value of $f(\beta, C)$ but where this particular element is replaced by the value d . If we apply the mean value theorem to this function we get:

$$|f(\beta, C, d_1) - f(\beta, C, d_2)| \leq \max_{d \in [d_1, d_2]} \left\{ \frac{\partial f(\beta, C, d)}{\partial d} \right\} |d_1 - d_2| \quad (\text{EC.2})$$

and

$$|p^*(\beta, C, d_1) - p^*(\beta, C, d_2)| \leq \max_{d \in [d_1, d_2]} \left\{ \frac{\partial p^*(\beta, C, d)}{\partial d} \right\} |d_1 - d_2|. \quad (\text{EC.3})$$

This implies that if $\|C - C'\|_\infty \leq c_1 \|\beta - \beta'\|_\infty$, for some c_1 then there are c_2 and c_3 such that $|p^*(\beta, C) - p^*(\beta', C')| \leq c_2 \|\beta - \beta'\|_\infty$ and $|f(\beta, C) - f(\beta', C')| \leq c_3 \|\beta - \beta'\|_\infty$.

We will now apply the above in combination with induction to our problem.

For $t = 0$,

$$|p^*(q_1, q_2, 1, \beta) - p^*(q_1, q_2, 1, \beta')| = |p^*(\beta, 0) - p^*(\beta', 0)| \leq \max_{i, j \in \{1, 2\} \times \{1, 2, O\}} \max_{b \in [\beta_{i,j}, \beta'_{i,j}]} \left\{ \frac{\partial p^*(\beta, 0, b)}{\partial \beta_{i,j}} \right\} \|\beta - \beta'\|_\infty$$

and

$$|G^*(q_1, q_2, 1, \beta) - G^*(q_1, q_2, 1, \beta')| = |f(\beta, 0) - f(\beta', 0)| \leq \max_{i,j \in \{1,2\} \times \{1,2,O\}} \max_{b \in [\beta_{i,j}, \beta'_{i,j}]} \left\{ \frac{\partial f(\beta, 0, b)}{\partial \beta_{i,j}} \right\} \|\beta - \beta'\|_\infty.$$

This implies that for some c

$$\begin{aligned} & |\Delta_i G^*(q_1, q_2, 0, \beta) - \Delta_i G^*(q_1, q_2, 0, \beta')| \\ & \leq |G^*(q_1, q_2, 0, \beta) - G^*(q_1, q_2, 0, \beta')| + |G^*((q_1, q_2) - \underline{e}_i, 0, \beta) - G^*((q_1, q_2) - \underline{e}_i, 0, \beta')| \\ & \leq c \|\beta - \beta'\|_\infty. \end{aligned}$$

For any $t > 0$, we assume the statement is true up to $t - 1$. We define

$$C := (\Delta_1 G^*(q_1, q_2, t - 1, \beta), \Delta_2 G^*(q_1, q_2, t - 1, \beta), \Delta_O G^*(q_1, q_2, t - 1, \beta)),$$

and

$$C' := (\Delta_1 G^*(q_1, q_2, t - 1, \beta'), \Delta_2 G^*(q_1, q_2, t - 1, \beta'), \Delta_O G^*(q_1, q_2, t - 1, \beta')),$$

such that

$$|p^*(q_1, q_2, t, \beta) - p^*(q_1, q_2, t, \beta')| = |p^*(\beta, C) - p^*(\beta', C')| \leq c_1 \|\beta - \beta'\|_\infty$$

and

$$|G^*(q_1, q_2, t, \beta) - G^*(q_1, q_2, t, \beta')| = |f^*(\beta, C) - f^*(\beta', C')| \leq c_2 \|\beta - \beta'\|_\infty,$$

for some constants c_1 and c_2 , because of (EC.2) and (EC.3) and the induction hypothesis. Note that the constants c_1 and c_2 continuously depend on β and β' , but as β is only defined on a closed and bounded spaces, the maximum of the constants is finite. $\square$

A LEMMA EC.2. *There exists a constant c such that for any $\beta' \in \mathcal{B}$*

$$G^* - G(p^*(\beta'), O^*) \leq c \|\beta - \beta'\|_\infty^2 \quad (\text{EC.4})$$

Proof For clarity, we denote p^* as $p^*(\beta)$, and G^* as $G(p^*(\beta), O^*)$ throughout the proof of this lemma. First, we observe that

$$\begin{aligned} & G(p^*(\beta), O^*) - G(p^*(\beta'), O^*) \\ & \leq G(p^*(\beta), O^*) - G(p^*(\beta), O^*, \beta') - [G(p^*(\beta'), O^*) - G(p^*(\beta'), O^*, \beta')], \end{aligned} \quad (\text{EC.5})$$

where $G(p, O, \beta')$ denotes the expected revenue obtained when the parameter matrix is β' and decisions p and O are applied. Hence, G^* can also be written as $G(p^*(\beta), O^*(\beta), \beta)$.

Then we define

$$\beta_s = \beta' + s(\beta - \beta'), \quad s \in [0, 1].$$

Because all the $h(\cdot)$ functions are twice continuous differentiable in β , G is differentiable in β . For any p ,

$$G(p, O^*, \beta) - G(p, O^*, \beta') = \int_0^1 \nabla_\beta G(p, O^*, \beta_s)^T (\beta - \beta') ds \quad (\text{EC.6})$$

and therefore, by applying equation EC.6 twice to inequality EC.5,

$$G(p^*(\beta'), O^*) - G(p^*(\beta'), O^*) \leq \int_0^1 [\nabla_\beta G(p^*(\beta), O^*, \beta_s) - \nabla_\beta G(p^*(\beta'), O^*, \beta_s)]^T (\beta - \beta') ds.$$

Because the $h_j(\cdot)$ are twice continuous differentiable and the parameter and decision spaces are bounded, there exist a constant c_1 , such that

$$|\nabla_\beta G(p^*(\beta), O^*, \beta_s) - \nabla_\beta G(p^*(\beta'), O^*, \beta_s)| \leq c_1 \|p^*(\beta) - p^*(\beta')\|_1,$$

and thus, using lemma EC.1, there exists constants c_2 and c such that

$$G(p^*(\beta), O^*) - G(p^*(\beta'), O^*, \beta) \leq c_2 \|p^*(\beta) - p^*(\beta')\|_1 \|\beta - \beta'\|_\infty \leq c \|\beta - \beta'\|_\infty^2.$$

□

B There exists a constant C such that for any p

LEMMA EC.3.

$$G^* - G(p, O^*) \leq c \max_{q_1 \leq Q_1, q_2 \leq Q_2, t \leq T} \{p^*(q_1, q_2, t) - p(q_1, q_2, t)\} \quad (\text{EC.7})$$

Proof For every state (q_1, q_2, t) , define

$$c_{q_1, q_2, t} := \max_{p \in [\underline{p}, \bar{p}]} \left| \frac{\partial G(q_1, q_2, t, p)}{\partial p} \right|.$$

Since $G(q_1, q_2, t, p)$ is continuously differentiable in p on the compact interval $[\underline{p}, \bar{p}]$, we have $c_{q_1, q_2, t} < \infty$. Since the state space and time horizon are finite, we may define

$$c := \max_{\substack{q_1 \leq Q_1, q_2 \leq Q_2 \\ t \leq T}} c_{q_1, q_2, t} < \infty.$$

Let $p^* = p^*(q_1, q_2, t)$. By the mean value theorem, for every feasible price $p = p(q_1, q_2, t)$, there exists some $\tilde{p}$ between p and p^* such that

$$\begin{aligned} G^*(q_1, q_2, t) - G(q_1, q_2, t, p) &= G(q_1, q_2, t, p^*) - G(q_1, q_2, t, p) \\ &= \frac{\partial G(q_1, q_2, t, \tilde{p})}{\partial p} (p^* - p). \end{aligned}$$

Consequently,

$$G^*(q_1, q_2, t) - G(q_1, q_2, t, p) \leq c |p^* - p|.$$

Hence,

$$\mathbb{E}[G^*(q_1, q_2, t) - G(q_1, q_2, t, p)] \leq c \mathbb{E}[|p^*(q_1, q_2, t) - p(q_1, q_2, t)|].$$

Applying this bound recursively over the finite horizon T yields

$$\mathbb{E}[G^* - G(Q_1, Q_2, T, p, O^*)] \leq c \max_{\substack{q_1 \leq Q_1, q_2 \leq Q_2 \\ t \leq T}} \mathbb{E}[|p^*(q_1, q_2, t) - p(q_1, q_2, t)|].$$

□

C LEMMA EC.4. *There exists a constant c such that for any decisions p , O and O'*

$$|G(p, O) - G(p, O')| \leq c \|O - O'\|_1^2 \quad (\text{EC.8})$$

Proof First we define the set

$$\mathcal{O}_k(O) := \{O' \text{ st. } \|O - O'\|_1 = k\}.$$

Then we define the maximum reward distance of all opaque strategies within this set, given some fixed price strategy as

$$d_k(p, O) := \max_{O'} \{|G(p, O) - G(p, O')| \text{ st. } O' \in \mathcal{O}_k(O)\}.$$

Now, we can complete the proof by choosing $c = \max_{p,k} \left\{ \frac{d_k(p, O)}{k^2} \right\}$.

D LEMMA EC.5. *There exists a constant c such that for any $\beta' \in \mathcal{B}$*

$$\|O^*(\beta) - O^*(\beta')\|_1 \leq c \|\beta - \beta'\|_\infty \quad (\text{EC.9})$$

Proof First, we mention that if $O^*(q_1, q_2, t, \beta) = \{1, 2\}$ or if $\Delta_1 G^*(q_1, q_2, t) = \Delta_2 G^*(q_1, q_2, t)$, that $O^*(q_1, q_2, t, \beta) = O^*(q_1, q_2, t, \beta')$ for all β' .

We define

$$d_\beta(s) := \min_{\beta'} \{\|\beta - \beta'\|_\infty \text{ st. } \|O^*(\beta) - O^*(\beta')\|_1 \leq s\} \text{ and } d_\beta := \max_{s \in \{1, \dots, (Q_1+1)(Q_2+1)T\}} \left\{ \frac{s}{d_\beta(s)} \right\}.$$

Note that $d_\beta < \infty$, because $G^*(q_1, q_2, t, \beta)$ is continuous in all elements of β . This means that, without loss of generality, in case $O^*(q_1, q_2, t, \beta) = 1$, there is an $\varepsilon > 0$, such that if $\|\beta - \beta'\|_\infty \leq \varepsilon$, $G^*(q_1 - 1, q_2, t, \beta') > G^*(q_1, q_2 - 1, t, \beta')$ and $O^*(q_1, q_2, t, \beta') = 1$.

Choosing $c = \max_\beta \{d_\beta\}$ completes the proof. □

EC.4. Proof of Theorem 4

THEOREM 4. Policy $\varphi_O(\varepsilon, \delta)$ is defined as follows:

$O(q_1, q_2, t, k) := O^*(q_1, q_2, t, \hat{\beta}_{k-1})$ and

1. If $\max\{|p_a - p_b| \text{ st. } p_a, p_b \in \{p_{T(k-1)+1}, \dots, p_{T_k-t}, p^*(q_1, q_2, t, k)\}\} < \varepsilon$ and $(t = 1 \text{ or } q_1 = 1 \text{ or } q_2 = 1)$ and $J_{k-1} < k^{\frac{1}{2}+\delta}$:

$$p(q_1, q_2, t, k) := \min\{p_{T(k-1)+1}, \dots, p_{T_k-t}\} + \varepsilon \text{ or}$$

$$p(q_1, q_2, t, k) := \max\{p_{T(k-1)+1}, \dots, p_{T_k-t}\} - \varepsilon$$

2. else if $\min\{p_{T(k-1)+1}, \dots, p_{T_k-t}, p^*(q_1, q_2, t, k)\} = \bar{p}$ and $(t = 2 \text{ or } q_1 = 2 \text{ or } q_2 = 2)$ and $J_{k-1} < k^{\frac{1}{2}+\delta}$:
 $p(q_1, q_2, t, k) := (\bar{p} + p)/2$

3. else:

$$p(q_1, q_2, t, k) := p^*(q_1, q_2, t, \hat{\beta}_{k-1}),$$

where $\hat{\beta}_{k-1}$ is the MLE based on the first $k-1$ repetitions, J_k is the number of the first k repetitions where $\max\{|p_a - p_b|, p_a, p_b \in \{p_{T(k-1)+1}, \dots, p_{T_k}\}\} > \varepsilon$. The order of the regret of policy $\varphi_O(\varepsilon, \delta)$ is

$$\mathbf{R}_k^{\varphi_O(\varepsilon, \delta)} = \begin{cases} O(k^{\frac{1}{2}+\delta}) & \text{if } \beta \in \mathbb{R}^{2 \times 3} \\ O(\log^2(k)) & \text{if } \beta \in \mathcal{E}^\varepsilon \end{cases}.$$

Proof of Theorem 4 First, we note that for any $\beta \in \mathbb{R}^{2 \times 3}$, this policy results in a $g_2(k)$ of $O(k^{\frac{1}{2}+\delta})$ and a $g_1(k) = O(\log(k)/k^{\frac{1}{2}+\delta})$, where $g_1(\cdot), g_2(\cdot)$ as in Theorem 3. The policy guarantees that out of the last k repetitions at least in $k^{\frac{1}{2}+\delta}$ repetitions there are two opaque prices with at least distance ε , what guarantees $g_1(k) = O(\log(k)/k^{\frac{1}{2}+\delta})$ because of equation (6). At the same time, the policy deviates at most $2\bar{p}$ from the perceived optimal price in $k^{\frac{1}{2}+\delta}$ out of k iterations, which results in $\sum_{l=1}^k \mathbb{E}[\|p^*(\hat{\beta}_l) - p_l\|_1] \leq 2\bar{p}k^{\frac{1}{2}+\delta}$. Together, this implies

$$\mathbf{R}_k^{\varphi_O} = O\left(\max\left\{\sum_{l=1}^k \log(l)/k^{\frac{1}{2}+\delta}, k^{\frac{1}{2}+\delta}\right\}\right) = O(k^{\frac{1}{2}+\delta}).$$

Second, if $\beta \in \mathcal{E}^\varepsilon$ we can bound the expected number of times active exploration is needed by a constant what results in $g_2(k) \leq c$, for some constant c . As argued before, we can obtain a $\rho(\varepsilon)$ such that $\|\beta - \beta'\|_1 \leq \rho(\varepsilon)$ implies $\beta' \in \mathcal{E}^\varepsilon$ and, using (3), an $T_{\rho(\varepsilon)}$, such that if $k > T_{\rho(\varepsilon)}$, $\hat{\beta}_k \in \mathcal{E}^\varepsilon$. Because $\{\beta' \text{ st. } \|\beta - \beta'\|_1 \leq \rho(\varepsilon)\}$ is closed and bounded and we consider an MDP with a finite amount of states, we can define q as the minimum probability over $\{\beta' \text{ st. } \|\beta - \beta'\|_\infty \leq \rho(\varepsilon)\}$ that within a repetition after $T_{\rho(\varepsilon)}$ two optimal opaque prices occur with distance $> \varepsilon$. Therefore, the probability given $k - x \geq T_{\rho(\varepsilon)}$, that within the last x repetitions, say $k - x, \dots, k$, repetitions this did not happen is therefore smaller than $(1 - q)^x$. The probability that in repetition k active learning is needed is the probability that the last $k^{\frac{1}{2}+\delta}$ repetitions no opaque prices occurred with difference at least ε . Since the policy deviates from the perceived optimal price only when active exploration is needed, and this deviation is bounded by $2\bar{p}$, we have

$$\begin{aligned} & \sum_{l=1}^k \mathbb{E}[\|p^*(\hat{\beta}_l) - p_l\|_1] \\ & \leq c_1 \mathbb{E}[T_{\rho(\varepsilon)}] + c_2 \sum_{l=1}^k (1 - q)^{l^{\frac{1}{2}+\delta}} \\ & \leq c_3. \end{aligned}$$

Together with (5) this implies that if $\beta \in \mathcal{E}^\varepsilon$,

$$\mathbf{R}_k^{\varphi O} = O \left(\max \left\{ \sum_{l=1}^k \log(l)/l, 1 \right\} \right) = O(\log^2(k)). \quad \square$$

EC.5. Detailed description of exploration of numerical results

In Figure EC.1 and EC.2 we present the average effect of the different parameters on expected revenue in waste. Next we present the top five of cases that showed the most and least effect in terms of absolute and relative difference in revenue and waste of the optimal policy and static policy. The top and bottom fives can be found in Table EC.1 and EC.2.

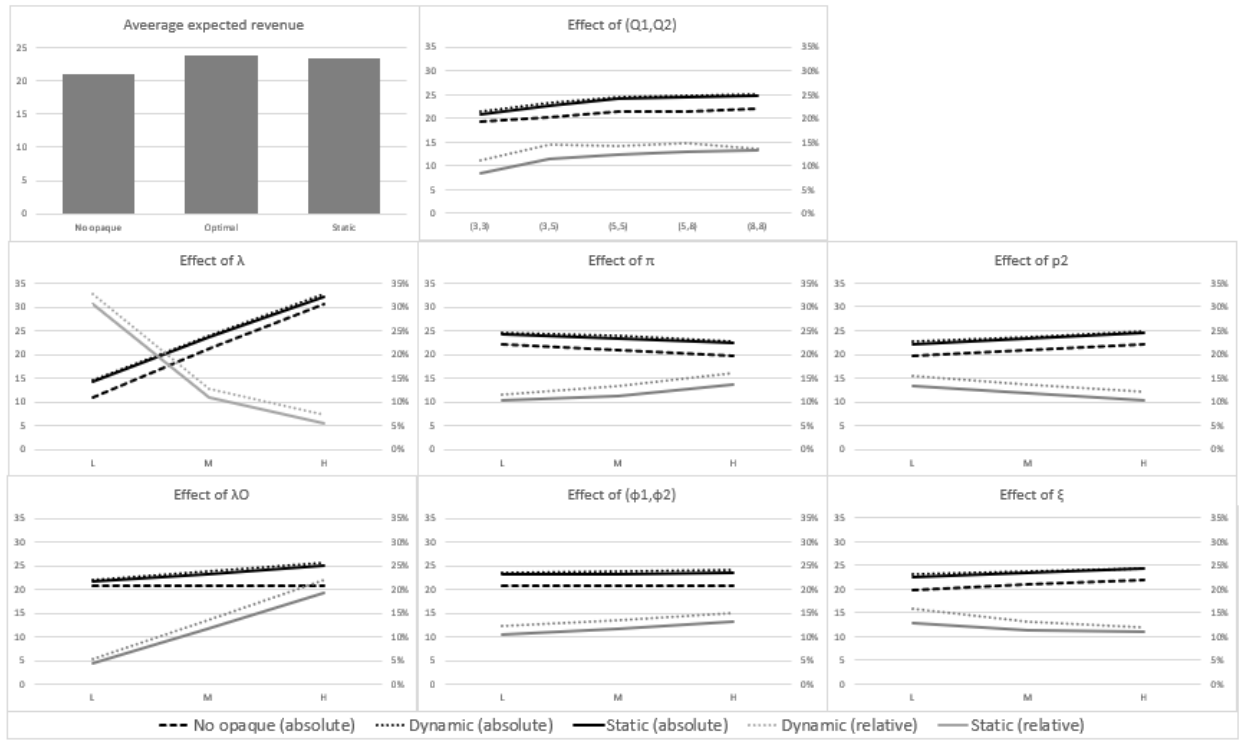
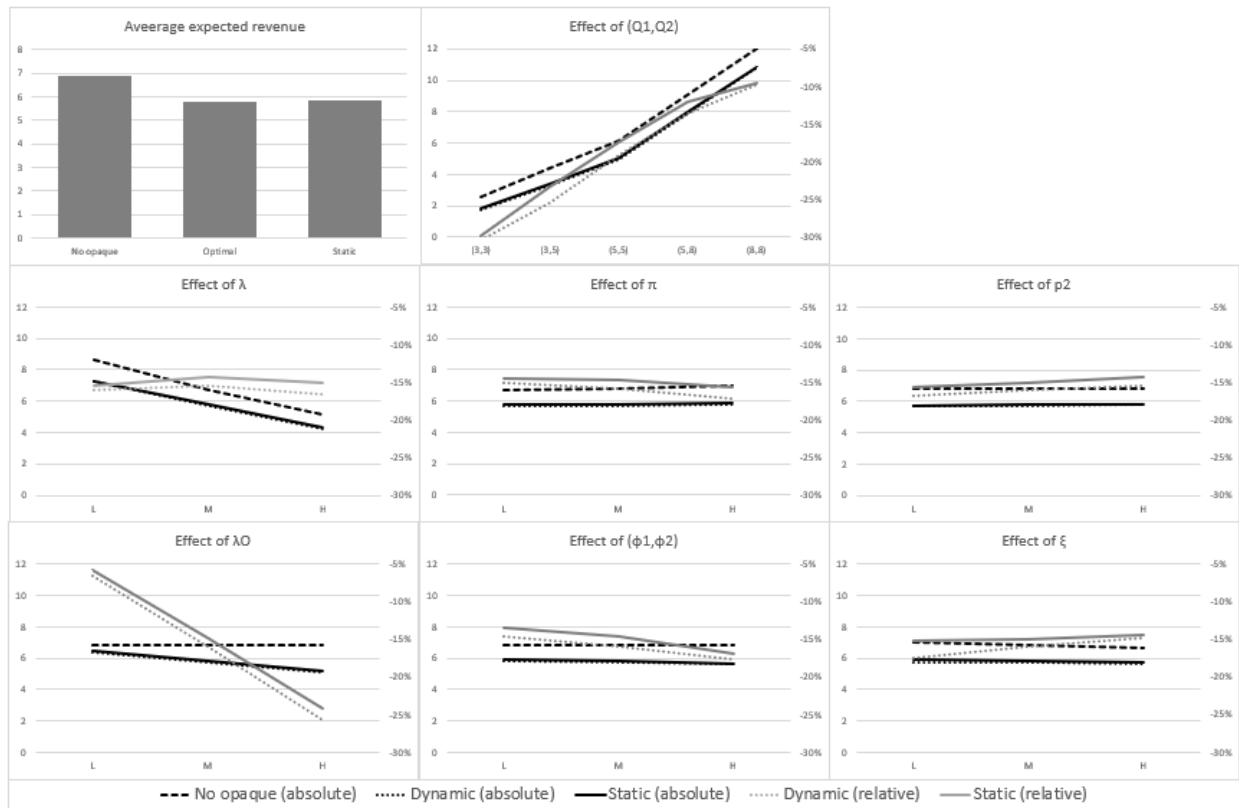


Figure EC.1 Main effects on expected revenue

**Figure EC.2** Main effects on expected revenue

Q_1	Q_2	λ	λ_O	ϕ_1	π_1	p_2	ζ	Expected revenue N O		Relative difference	Absolute difference
Top and bottom five Absolute difference											
3	5	0.3	0.3	2	0.9	5.2	0	26.00	18.30	42%	7.70
3	5	0.3	0.3	2	0.9	6.2	0	26.47	18.90	40%	7.57
3	5	0.3	0.3	1	0.9	5.2	0	25.83	18.30	41%	7.53
3	5	0.3	0.3	2	0.9	7.2	0	26.93	19.50	38%	7.43
3	5	0.3	0.3	1	0.9	6.2	0	26.21	18.90	39%	7.32
3	3	0.3	0.1	1	0.7	7.2	0.5	29.85	29.77	0%	0.08
3	3	0.3	0.1	1	0.5	7.2	0.5	31.23	31.14	0%	0.09
3	5	0.3	0.1	1	0.5	7.2	0.5	35.40	35.30	0%	0.10
3	3	0.3	0.1	1	0.5	7.2	0.99	32.31	32.19	0%	0.11
8	8	0.3	0.1	1	0.5	7.2	0.5	37.31	37.19	0%	0.12
Top and bottom five Relative difference											
3	5	0.1	0.3	3	0.9	5.2	0	16.27	9.69	68%	6.58
3	5	0.1	0.3	3	0.9	6.2	0	16.42	9.89	66%	6.53
5	8	0.1	0.3	3	0.9	5.2	0	17.06	10.36	65%	6.70
5	5	0.1	0.3	3	0.9	5.2	0	17.04	10.36	64%	6.68
5	8	0.1	0.3	3	0.9	5.2	0.5	17.07	10.38	64%	6.69
3	3	0.3	0.1	1	0.7	7.2	0.5	29.85	29.77	0%	0.08
3	5	0.3	0.1	1	0.5	7.2	0.5	35.40	35.30	0%	0.10
3	3	0.3	0.1	1	0.5	7.2	0.5	31.23	31.14	0%	0.09
8	8	0.3	0.1	1	0.5	7.2	0.5	37.31	37.19	0%	0.12
8	8	0.3	0.1	1	0.5	7.2	0.99	37.32	37.20	0%	0.12

Table EC.1 Top and bottom five settings with respect to different revenue measures

Q_1	Q_2	λ	λ_O	ϕ_1	π_1	p_2	ζ	Expected waste N O		Relative difference	Absolute difference
Top and bottom five Absolute difference											
3	5	0.3	0.3	2	0.9	5.2	0	4.48	1.66	153%	2.82
3	5	0.3	0.3	2	0.9	6.2	0	4.48	1.69	153%	2.79
3	5	0.3	0.3	1	0.9	5.2	0	4.48	1.71	153%	2.77
3	5	0.3	0.3	2	0.9	7.2	0	4.48	1.73	153%	2.75
3	5	0.3	0.3	1	0.9	6.2	0	4.48	1.75	153%	2.73
3	3	0.3	0.1	1	0.5	7.2	0.5	0.98	0.93	3%	0.05
3	3	0.3	0.1	1	0.5	7.2	0.99	0.81	0.76	15%	0.05
3	3	0.3	0.1	1	0.9	6.2	0.99	0.81	0.76	12%	0.06
3	3	0.3	0.1	1	0.5	6.2	0.99	0.81	0.75	13%	0.06
3	3	0.3	0.1	1	0.7	6.2	0.99	0.81	0.75	13%	0.06
Top and bottom five Relative difference											
3	3	0.3	0.3	2	0.9	5.2	0	2.48	0.81	140%	1.68
3	3	0.3	0.3	1	0.9	5.2	0	2.48	0.83	140%	1.66
3	3	0.3	0.3	2	0.9	6.2	0	2.48	0.84	140%	1.64
3	3	0.3	0.3	1	0.9	6.2	0	2.48	0.86	140%	1.62
3	3	0.3	0.3	2	0.9	7.2	0	2.48	0.87	140%	1.61
8	8	0.3	0.1	1	0.5	7.2	0.5	10.00	9.89	1%	0.11
8	8	0.3	0.1	1	0.5	7.2	0.99	10.00	9.89	1%	0.11
8	8	0.3	0.1	1	0.5	7.2	0	10.00	9.89	1%	0.11
8	8	0.3	0.1	1	0.7	7.2	0.99	10.00	9.83	0%	0.17
8	8	0.3	0.1	1	0.7	7.2	0.5	10.01	9.84	2%	0.17

Table EC.2 Top and bottom five settings with respect to different waste measures