Optimal Pricing and Inventory Control Policy with Quantity-Based Price Differentiation

Ye Lu, Youhua (Frank) Chen, Miao Song, Xiaoming Yan

Abstract

A firm facing price dependent stochastic demand aims to maximize its total expected profit over a planning horizon. In addition to the regular unit selling price, the firm can utilize quantity discounts to increase sales. We refer to this dual-pricing strategy as quantity-based price differentiation. At the beginning of each period, the firm needs to make three decisions: replenish the inventory, set the unit selling price if the unit sales mode is deployed, and set the quantity-discount price if the quantity-sales mode is deployed (or the combination of the two modes of sales). We identify conditions under which the optimal inventory control policy and selling/pricing strategy is well-structured. Remarkably, under a utility-based demand framework, these conditions can be unified by a simple regularity assumption that has long been used in the auction and mechanism design literature. Moreover, sharper structural results are yielded for the optimal selling strategy. We also examine the comparative advantage of quantity-based price differentiation with respect to model parameters. Our numerical study shows that substantial profit improvement can be gained as a result of shifting from uniform pricing to quantity-based pricing, especially when the product has a low unit ordering cost and high utility.

1 Introduction

1.1 Background and an Overview of the Main Findings

Customers often see advertisements proclaiming “buy one and get a second for 50% off” and so on. For example, in early 2011, DrugStore.com promoted filters under the banner “Buy one 3M filter and get a second for 50% off”. Similar deals are also commonly advertised, such as “4-for-3” for garden and glassware products at Amazon.com and “buy 2 get 1 free” on selected brands at clickinks.com, an online distributor of printer ink and toner cartridges. These examples represent a simple yet common type of quantity discount program used by retailers whereby consumers are given the choice of buying one unit of a product at the full price or buying up more units to obtain a discount. We refer to the sale of units at their full (or regular) price as unit sales (or the unit-sales mode) and the sale of units at a discount as quantity sales (or the quantity-sales mode).

In the past decade or so, dynamic pricing has become increasingly accessible as a useful tool for retailers and manufacturers to better match supply with demand and increase profit. From the perspective of the retailer, the quantity-sales mode is an additional tool to dynamic pricing. We refer to standard dynamic pricing as (the strategy of) uniform pricing (or uniform pricing in short) and a combination of pricing and quantity-based selling as (the strategy of) quantity-based pricing (or quantity-based pricing in short). Compared with uniform pricing, by which a single price is charged for any quantity purchased by customers, quantity-based pricing
extends the unit-sales mode by introducing the quantity-sales mode, in which two prices may be offered, one for the unit-sales mode and the other for the quantity-sales mode. This differentiated dynamic pricing strategy is common in retailing, especially online retailing. For example, between March 3, 2013 and March 15, 2013, the selling mode and prices of a particular product at the online store of ParknShop, a major supermarket chain in Hong Kong, evolved as shown in Table 1.

Table 1: C&S Greenie 4 Ply Toilet Roll

<table>
<thead>
<tr>
<th>Date</th>
<th>Modes</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 3</td>
<td>unit only</td>
<td>$31.9</td>
</tr>
<tr>
<td>March 8</td>
<td>unit only</td>
<td>$28.9</td>
</tr>
<tr>
<td>March 11</td>
<td>unit and quantity</td>
<td>$28.5, buy 2 to save $5</td>
</tr>
<tr>
<td>March 15</td>
<td>unit and quantity</td>
<td>$30.5, buy 2 to save $9</td>
</tr>
</tbody>
</table>

In this paper, we study the interplay between quantity-based pricing and inventory control. Demand is stochastic and price-dependent. Replenishment orders can be placed in any period and incur linear purchasing costs. In each period, the firm makes three decisions: the order-up-to inventory level of the product, the unit-sales price, and the quantity-sales price for a given number of multiple units. Demand, which depends on both prices, then arrives and is fulfilled by on-hand inventories. Finally, inventory holding and shortage costs are evaluated. The decisions in each period include: (1) the inventory replenishment quantity and (2) the selling mode(s) and the respective price(s). The objective is to maximize the total expected profit over a finite or infinite horizon. In particular, we attempt to address three questions. First, in a dynamic setting with inventory replenishment, what are the optimal inventory control policy and the optimal selling mode(s)? More specifically, when is it beneficial to engage simultaneously in the unit- and quantity-sales modes (for example, “buy one for $20 and two for $30”), or one of the two selling modes? Second, if the two selling modes are offered simultaneously, customers can save some money by buying $m$ units in the quantity sales mode. We name this amount of saving as the quantity-discount amount. How does the quantity-discount amount depend on the optimal inventory level? Third, the retailer may obtain additional profit by offering the quantity-sales mode. How does the benefit of quantity-based price differentiation depend on the system parameters? Moreover, under what conditions could quantity-based price differentiation bring significant value to the firm in comparison to uniform pricing strategy where the firm charges a uniform price (yet possibly time-varying) in each period?

Before reviewing the literature and putting our contributions into perspective, we highlight the main results pertaining to the aforementioned questions. In Section 2, we propose a demand model which is based on a three-choice framework (the unit-sales only mode, the quantity-sales only mode, or a combination of the two), and identify the conditions under which the inventory-pricing problem has a well-structured inventory control and selling strategy. We find that the optimal inventory decision follows a base-stock policy, and the optimal selling strategy as illustrated by Figure 1 depends on the optimal order-up-to level $y^*_t$ in period $t$ ($y^*_t$ is equal to the base-stock level if the initial inventory level is lower than the base-stock level, and is equal to the initial inventory level if the initial inventory level is no less than the base-stock level). Figure 1 implies that as the
inventory level increases, the quantity-sales mode becomes the more likely option and the unit-sales mode becomes less likely. Although such a structural result is intuitive, if one or more of the model assumptions are unmet, the optimal structure no longer follows what Figure 1 depicts. This optimal structure also answers the question of when one or both of the two sales modes should be adopted.

Figure 1: Optimal Pricing Policy for the Optimal Order-up-to Level $y_t^*$

In Section 3, we use a stochastic utility (or valuation) model to characterize the demand. Remarkably, we show that if, and only if the customers’ virtual value function is increasing, which is equivalent to a simple regularity assumption in the auction and mechanism design literature (see Myerson 1981), all the assumptions in Section 2 hold true, and hence the optimal inventory-selling strategy follows Figure 1. (In the current context, the virtual value function measures the sensitivity of revenue to customer demand, see Section 3.) We also reveal the following properties of the optimal selling/pricing strategy. (i) As long as inventory replenishment is required, it is never optimal to use the quantity-sales mode alone. This implies that under a steady environment, it is never optimal to sell only in the quantity-sales mode. (ii) Let $p_1^*$ be the optimal unit selling price and $p_2^*$ be the optimal price for every $m$ units. Then, the quantity discount, $mp_1^* - p_2^*$, may be increasing or decreasing with the inventory level. It is counter-intuitive that the discount amount is decreasing with the inventory level. The direction of monotonicity hinges on how the virtual value function varies with the change in the unit-selling price. (iii) Although the amount of the quantity discount does not always increase or decrease with the inventory level, it is shown that the optimal prices $p_1^*$ and $p_2^*$ both decrease with the inventory level. Moreover, $p_2^* - p_1^*$ decreases with the inventory level, i.e., $p_2^*$ always decreases faster than $p_1^*$. Therefore, Section 3 answers the second research question.

The value of quantity-based pricing is defined as the profit difference between two systems with/without the quantity-sales mode. In Section 4, we show that this value decreases with the ordering cost but increases with the incremental utility (value) that the additional $(m - 1)$ units of product generate for customers. These results partially answer the first and third research questions in terms of the system parameters. Finally, we conduct a numerical study in Section 4 to see the relative magnitudes of profit gain when shifting from a static pricing strategy to a dynamic pricing strategy (i.e., uniform pricing) and then to a quantity-based pricing strategy. Our numerical analysis demonstrates that quantity-based pricing can bring significant additional value for a large range of values for each parameter, especially for cases with a long planning horizon, low unit ordering cost, and high product utility. Our numerical studies also show that shifting from a dynamic pricing
strategy (or uniform pricing) to a quantity-based pricing strategy often improves profitability much more than shifting from a fixed pricing strategy to a dynamic pricing strategy. Therefore, Section 4 answers the third research question.

1.2 Literature Review and the Positioning of the Current Work

Numerous economics and marketing studies have explained how quantity-based price differentiation can be used to improve a firm’s profit. In economics terms, offering quantity discounts is a form of second-degree price discrimination. From a marketing point of view, quantity discounts afford manufacturers the ability to price discriminate between high-volume and low-volume users (see Dolan 1987). From a demand perspective, the rationale for quantity discounts arises from decreasing consumers’ marginal willingness-to-pay (WTP).

A stream of the operations management literature has investigated the coordination of inventory control (replenishment) and quantity discounts. Research in this field can be split into two lines: that from the buyer’s perspective and that from the seller’s perspective. The former focuses on the problem of determining the economic order quantities for a buyer who is offered a quantity discount. Studies on the buyer’s perspective include Subramanyam and Kumaraswamy (1981), Rubin, Dilts and Barren (1983), Sethi (1984), and Jucker and Rosenblatt (1985). In contrast, Monahan (1984), Lal and Staelin (1984), and Lee and Rosenblatt (1986) study the problem from the seller’s perspective. They seek to derive the optimal quantity-based pricing schemes that maximize the seller’s profit. Either way, the two streams of research conclude that quantity discounts can improve the profit of either the buyer or the seller, and sometimes that of the whole supply chain. However, these studies deal with deterministic demand or single period models. For a general review of the literature on quantity discounts, we refer readers to Dolan (1987) and Munson and Rosenblatt (1998). Our model extends the seller’s model to allow demand uncertainty in a multi-period setting.

Porteus (1971), among others, studies the buyer’s model with stochastic demand and concave ordering costs. He shows that under certain conditions, a generalized \((s, S)\) policy is optimal in a finite-horizon setting. Altintas, Erhun and Tayur (2008; AET hereafter) were the first to study quantity discounts under stochastic demand for multiple periods from the seller’s perspective. Our model differs from the AET model in the following significant ways. First, AET deal with transactions between a supplier and a customer, whereas we consider a firm selling to the market. Hence, in our model, the inventory decision and pricing decision are centralized to the selling firm, while in the AET model, the inventory decision and pricing decision are decoupled to the buyer and the seller, who have different objectives. Second, in our model, the seller adopts a dynamic pricing strategy while in the AET model, the quantity discount scheme once decided in the first period is fixed in subsequent periods.

Our paper is closely related to the growing body of research on inventory and pricing coordination. Federgruen and Heching (1999) study the situation in which a retailer who faces stochastic and price-dependent demand dynamically makes pricing and inventory decisions to maximize her total expected profit over a finite/infinite horizon. A number of further developments have since been published (e.g., for inventory-pricing problems, see Chen and Simichi-Levi 2004a,b; Huh and Janakiraman 2008; and Song et al. 2009). For an
earlier literature review, see Yano and Gilbert (2003). Chen and Simchi-Levi (2012) provide an up-to-date survey of the inventory-pricing literature. In these papers, a uniform price is charged regardless of the number of purchased units. In addition to two decisions, the order quantity and per-unit price, that are made in each period in the existing inventory-pricing models, our model adds one more decision, quantity-based pricing.

The revenue management literature also investigates dynamic pricing decisions. However, these studies typically do not allow for inventory replenishment and quantity-based price differentiation, whereas our paper explicitly considers the interplay of inventory replenishment and quantity-based pricing decisions. See Bitran and Caldentey (2003) and Elmaghraby and Keskinocak (2003) for surveys of the revenue management and dynamic pricing literature.

In view of the literature, our contribution is two-fold. First, this paper is among the first to model the quantity-based pricing strategy in a stochastic, periodic-review, inventory setting and address the aforementioned three new research questions, which do not exist in the traditional models with the unit-sales mode only. Second, on the theoretical side, we show that three assumptions we make on the demand model can be unified by the simple regularity assumption (the increasing virtual value function) that has long been identified in the auction and mechanism design literature. This remarkable consistency reveals the general applicability of the virtual value function.

All the proofs of the results in this paper can be found in Appendix A.

2 Model and Analysis

2.1 Formulation

Consider a firm that sells a single product over a horizon of \( n \) periods. To streamline the notation, we consider a system with stationary parameters over time. Nevertheless, the results derived here can easily be extended to a nonstationary setting. At the beginning of each period \( t = 1, \ldots, n \), the firm can place an order after observing the initial inventory level, \( x_t \). An order incurs a constant per-unit cost, \( c \). Assuming zero lead-time, the inventory level is raised instantaneously to \( y_t \geq x_t \). At the same time, the firm can quote up to two prices to sell the product: price \( p_{1t} \) for the unit-sales mode and price \( p_{2t} \) for the quantity-sales mode, which consists of (pre-fixed) \( m \) units. (For example, in the “buy two get one free” promotion, \( p_{2t} = 2p_{1t} \) and \( m = 3 \).) Demand for the product is then realized. At the end of period \( t \), the unsatisfied demand (leftover inventory) is backlogged (carried over) to the next period. (Discussion of relaxing the assumptions of zero lead-time, zero fixed-order cost, and including only two selling modes is included in Appendix B.)

In each period \( t \), we introduce a non-negative random variable \( D_t \), which we call the market size. Let \( p_t = (p_{1t}, p_{2t}) \). Then, demand in period \( t \), \( \xi_t \), is expressed as

\[
\xi_t = [\lambda_1(p_t) + m\lambda_2(p_t)]D_t + \varepsilon_t, \quad t = 1, \ldots, n, \tag{2.1}
\]

where \( \lambda_1(p_t) \) is the market share of the unit-sales mode, and \( \lambda_2(p_t) \) is the market share of the quantity-sales mode. Facing one or two selling modes with respective prices, a customer may choose to buy nothing. Then,
$\lambda_1 + \lambda_2 \leq 1$, and $\lambda_1 + \lambda_2$ can be strictly less than 1 because of the no-buy option. We assume $D_t$ to be independently and identically distributed over time and let $\mu_t = E[D_t] > 0$.

As both modes are basically selling the same product, their demand relies first on the number of customers who want this product and then on how customers split their demand among the two modes. Therefore, the demands for unit and quantity sales must have some strong correlation, which is captured by the first part of this demand model $[\lambda_1(p_t) + m\lambda_2(p_t)]D_t$. However, if we only consider this part – without the random term $\varepsilon_t$, then the demand for unit-sales $\lambda_1(p_t)D_t$ and the demand for quantity-sales $m\lambda_2(p_t)D_t$ are always perfectly positively correlated, which is an extreme case.

Other factors besides price can also affect demand, such as weather and social events. For example, the quantity of a soft drink sold through the quantity-sales mode may increase more than that sold through the unit-sales mode when the weather in summers becomes hot. Moreover, some customers are not price-sensitive. Fluctuations in the number of price-insensitive customers can also cause overall demand to vary. To capture those additional demand uncertainties, we include in the demand model (2.1) a random term $\varepsilon_t$. Here, we can assume $\varepsilon_t = \varepsilon_{1t} + m\varepsilon_{2t}$ where $\varepsilon_{1t}$ and $\varepsilon_{2t}$ are random noises with zero means for the unit-sales mode and the quantity-sales mode respectively. Note the demand for unit-sales $\lambda_1(p_t)D_t + \varepsilon_{1t}$ and the demand for quantity-sales $m(\lambda_2(p_t)D_t + \varepsilon_{2t})$ are not perfectly correlated any more because we do not make any assumption on the correlation between $\varepsilon_{1t}$ and $\varepsilon_{2t}$.

The model (2.1), which is called the additive-multiplicative model, was first proposed by Young (1978) and has been widely adopted in the pricing literature; see Chen and Simchi-Levi (2004a,b), Kocabiyikoglu and Popescu (2011), Lu and Sinchi-Levi (2013), and Roels (2013). Feng et al. (2013) also develop an estimation-optimization framework that can estimate this type of demand model using real data and optimally determine the price-inventory decisions. The model (2.1) can be reduced to a multiplicative demand model by setting $\varepsilon_t \equiv 0$. It can also be reduced to an additive demand model by setting $D_t$ as a constant.

For notational conciseness, we drop the subscript $t$ in the prices. For regular products, it is innocuous to assume that $\lambda_i(p_1,p_2)$ decreases in $p_i$ and increases in $p_j$, where $i = 1$ or 2, and $j = 3 - i$.

Following a commonly adopted technique, we consider the market shares instead of prices as the decision variables. Define

$$\mathcal{A} = \{ (\lambda_1, \lambda_2) : \lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_1 + \lambda_2 \leq 1 \},$$

which is the possible set of market shares. Similar to Song and Xue (2007), we assume a one-to-one correspondence between the price vector $p = (p_1, p_2)$ and market shares $(\lambda_1, \lambda_2)$. Let $(p_1(\lambda_1, \lambda_2), p_2(\lambda_1, \lambda_2))$ be the inverse functions of $(\lambda_1(p_1, p_2), \lambda_2(p_1, p_2))$. Hence, the feasible set for $(p_1, p_2)$ is

$$\mathcal{P} = \{ (p_1(\lambda_1, \lambda_2), p_2(\lambda_1, \lambda_2)) : (\lambda_1, \lambda_2) \in \mathcal{A} \}.$$  

**Remark 2.1.** To substantiate the demand model (2.1)-(2.3), we now consider an example of valuation-based demand model for illustration. Let $V$ denote the value (i.e., utility) of one unit of the product. Suppose $V$ to be a nonnegative random variable with a cumulative distribution function $G(p)$ and support $[0, \pi]$. We assume
that $G(p)$ is strictly increasing in $p$, and hence the inverse function of $G(p)$, $G^{-1}(x)$, is well defined for any $x \in [0, 1]$. Moreover, the value of $m$ units of the product is assumed to be $aV$, where $a$ is a constant with $a \in (1, m)$, reflecting the notion of a decreasing marginal utility in quantity. Therefore, the market shares for the unit sales mode at $p_1$ and for the quantity-discount sales mode at $p_2$, respectively, can be derived as follows

$$
\lambda_1(p_1, p_2) = \Pr(V - p_1 > aV - p_2, V - p_1 > 0) = \Pr(p_1 < V < \frac{p_2 - p_1}{a - 1}),
$$
$$
\lambda_2(p_1, p_2) = \Pr(aV - p_2 \geq v - p_1, aV - p_2 > 0) = \Pr(V \geq \frac{p_2 - p_1}{a - 1}, V > \frac{p_2}{a}).
$$

(2.4)

Note that the constraint $V - p_1 > 0$ (or $aV - p_2 > 0$) may be called the participation constraint because otherwise the buyer will not purchase, whereas the constraint $V - p_1 > aV - p_2$ (or $aV - p_2 \geq V - p_1$), the incentive compatibility constraint, ensures that the buyer has an incentive to buy one unit (or $m$ units).

According to (2.4), for any $p_1 \in [0, \overline{V}]$, $\lambda_1(p_1, p_2) = 0$ if and only if $ap_1 \geq p_2$. Hence, we only need to consider the region $ap_1 \leq p_2$, which implies $\frac{p_2 - p_1}{a} \geq \frac{p_2}{a}$. As a result, the market share functions in (2.4) can be rewritten as

$$
\lambda_1(p_1, p_2) = \Pr(p_1 < v < \frac{p_2 - p_1}{a - 1}) = G\left(\frac{p_2 - p_1}{a - 1}\right) - G(p_1),
$$
$$
\lambda_2(p_1, p_2) = \Pr(v \geq \frac{p_2 - p_1}{a - 1}) = 1 - G\left(\frac{p_2 - p_1}{a - 1}\right).
$$

(2.5)

Hence, the inverse functions of $\lambda_1(p_1, p_2)$ and $\lambda_2(p_1, p_2)$ follow

$$
p_1(\lambda_1, \lambda_2) = G^{-1}(1 - \lambda_1 - \lambda_2) \quad \text{and} \quad p_2(\lambda_1, \lambda_2) = (a - 1)G^{-1}(1 - \lambda_2) + G^{-1}(1 - \lambda_1 - \lambda_2),
$$

(2.6)

and the feasible set of $(p_1, p_2)$ is

$$
\mathcal{P} = \{(p_1, p_2) : 0 \leq p_1 \leq \overline{V}, \; ap_1 \leq p_2 \leq (a - 1)\overline{V} + p_1\}.
$$

(2.7)

This demand model will be the focus of the next section.

In general, the expected one-period revenue is

$$
R(\lambda_1, \lambda_2) = \left[p_1(\lambda_1, \lambda_2)\lambda_1 + p_2(\lambda_1, \lambda_2)\lambda_2\right] \cdot \mu,
$$

(2.8)

for any $(\lambda_1, \lambda_2) \in \mathcal{P}$. Now, prices essentially disappear from the revenue function.

We let $\ell(\cdot)$ be the inventory (holding/backlogging) cost after demand realization in period $t$. We assume $\ell(\cdot)$ to be convex and $\lim_{|x| \to \infty} \ell(x) = \infty$. Then, the expected inventory cost in period $t$ is $E[\ell(y - \xi_t)]$, where demand $\xi_t$ is defined in (2.1).

Let $V_t(x)$ be the maximum expected discounted profit for periods $t, t + 1, \ldots, n$ at the beginning of period $t$ with the inventory level $x$, under a given discount factor $\theta \leq 1$. The functions $V_t(x)$ satisfy

$$
V_t(x) = \max_{y \geq x, (\lambda_1, \lambda_2) \in \mathcal{P}} \left\{-c(y - x) - E[\ell(y - \xi_t)] + R(\lambda_1, \lambda_2) + \theta E[V_{t+1}(y - \xi_t)]\right\}, \text{ for any } t = 1, \ldots, n,
$$

(2.9)

where $V_{n+1}(x)$ is assumed to be concave, e.g., $V_{n+1}(x) = -ex^-$ with $e > 0$ where $x^- = \max\{-x, 0\}$.

The current period ($t$) and its immediate next period ($t + 1$) interact through the inventory level $y$ and demand $\xi_t$. For this reason, the overall demand rate $\lambda = \lambda_1 + m\lambda_2$ plays an important role in such an
interaction, because the expected demand \( E(\xi_t) = \lambda E(D_t) = \lambda \mu \). As \( \mu \) is a constant, \( \lambda \) is the only decision that determines the expected demand. For this reason, we introduce a new set

\[
\mathcal{B} = \left\{ (\lambda, \lambda_2) : \lambda = \lambda_1 + m \lambda_2, \ (\lambda_1, \lambda_2) \in \mathcal{A} \right\}. \tag{2.10}
\]

Clearly, \( \mathcal{A} \) and \( \mathcal{B} \) have a one-to-one correspondence under a linear transformation. Hence, \( \mathcal{B} \) is a convex set for any \( t = 1, \ldots, n \) because linear transformation preserves convexity.

The expected revenue in each period can then be represented as a function of \( \lambda \) and \( \lambda_1 \) or \( \lambda_2 \), as follows. Let

\[
R_1(\lambda_1, \lambda) = R\left( \lambda_1, \frac{\lambda - \lambda_1}{m} \right),
\]

which represents the expected one-period revenue as a function of \( \lambda \) and \( \lambda_1 \), and

\[
R_2(\lambda, \lambda_2) = R(\lambda - m \lambda_2, \lambda_2),
\]

which represents the expected one-period revenue as a function of \( \lambda \) and \( \lambda_2 \). The three revenue functions will be instrumental for the subsequent structural analysis.

For any given overall demand rate \( \lambda \), we denote the corresponding feasible region of share \( \lambda_2 \) using

\[
\mathcal{B}(\lambda) = \left\{ \lambda_2 : (\lambda, \lambda_2) \in \mathcal{B} \right\}. \tag{2.11}
\]

Note that \( \lambda \) affects the inventory dynamics, and \( \lambda_1 \) and \( \lambda_2 \) determine the current revenue. Let

\[
\lambda_2^*(\lambda) = \arg\max_{\lambda_2 \in \mathcal{B}(\lambda)} \left\{ R_2(\lambda, \lambda_2) \right\} \quad \text{and} \quad \Gamma(\lambda) = \max_{\lambda_2 \in \mathcal{B}(\lambda)} \left\{ R_2(\lambda, \lambda_2) \right\} = R_2(\lambda, \lambda_2^*(\lambda)), \tag{2.12}
\]

i.e., the function \( \lambda_2^*(\lambda) \) is defined as the value of share \( \lambda_2 \) that maximizes the expected revenue \( R_2(\lambda, \lambda_2) \) for a given \( \lambda \), and \( \Gamma(\lambda) \) is the corresponding maximum expected revenue.

To facilitate the analysis of the optimal policy, let \( W_t(x) = V_t(x) - cx \) and insert the definition of \( \Gamma(\lambda) \) into the optimality equation (2.9), to obtain

\[
W_t(x) = \max_{\lambda_2 \geq 0, \lambda_1 \geq m} \left\{ J_t(y, \lambda) \right\}, \tag{2.13}
\]

where

\[
J_t(y, \lambda) = -cy - L(y, \lambda) + \Gamma(\lambda) + \theta c \lambda \mu - \theta cy; \quad \text{the latter being the modified single-period loss function.} \tag{2.14}
\]

**Remark 2.2.** In the above formulation, we assumed zero cost in selling \( m \) units as a bundle. If this is not the case, for example, as one of the authors observed, in an airport shop where three boxes of chocolates are taped together for sale, suppose that bundling \( m \) units together costs \( c''m \). Then, the new revenue function can be obtained by modifying (2.8) minus the bundling costs: \( \hat{R}(\lambda_1, \lambda_2) = \left[ p_1(\lambda_1, \lambda_2) \lambda_1 + (p_2(\lambda_1, \lambda_2) - c''m) \lambda_2 \right] \cdot \mu \). Hence, \( \hat{R}(\lambda_1, \lambda_2) = R(\lambda_1, \lambda_2) - c''m \lambda_2 \mu \). As the difference between \( \hat{R}(\lambda_1, \lambda_2) \) and \( R(\lambda_1, \lambda_2) \) is a linear term, it does not change any of the analysis in the remainder of the paper. \( \square \)
2.2 Optimal Policy

The optimal policy comprises two decisions: inventory control and price setting. Price setting also determines the selling mode – a unit-sales price means a uniform price for any quantity sold and two prices means both unit-sales and quantity-sales prices. We first characterize the optimal inventory control policy and then determine the corresponding optimal selling/pricing pattern.

Intuitively, inventory should be replenished according to the base-stock policy. This is formally stated in Theorem 2.1 below. To prepare for the inventory policy, we first introduce one assumption for the one-period expected revenue function.

Assumption 2.1. Revenue function $R(\lambda_1, \lambda_2)$ is strictly joint concave in $(\lambda_1, \lambda_2)$ in any period.

This assumption is consistent with the principle of a diminishing marginal rate of return. It ensures that $R_2(\lambda, \lambda_2)$ is strictly concave in $(\lambda, \lambda_2)$ because it is a linear transformation of $R(\lambda_1, \lambda_2)$. Consequently, $\Gamma(\lambda)$ defined by (2.12) is strictly concave in $\lambda$ (see Proposition B-4 in Heyman and Sobel 1984).

For any given order-up-to level $y$, we also define $\lambda_t^*(y) = \arg\max_{0 \leq \lambda \leq m} \left\{ J_t(y, \lambda) \right\}$, where $J_t(y, \lambda)$ is defined in (2.14). Therefore, $\lambda_t^*(y)$ is the optimal overall demand rate achievable when the amount of inventory (after ordering) in period $t$ is $y$.

Theorem 2.1. (Inventory Policy) Let the optimal base-stock level

\[
S_t = \arg\max_{0 \leq \lambda \leq m} \left\{ J_t(y, \lambda) \right\} = \arg\max_{0 \leq \lambda \leq m} \left\{ \max_{0 \leq \lambda \leq m} \left\{ J_t(y, \lambda) \right\} \right\}.
\]  

(2.16)

If Assumption 2.1 holds, then the optimal inventory control policy is of the base-stock type; i.e., the optimal order-up-to level $y_t^* = S_t$ if the initial inventory level $x_t \leq S_t$ and $y_t^* = x_t$ otherwise, $t = 1, ..., n$.

The base-stock policy is derived in Theorem 2.1 without specifying when the firm should quote both prices, the unit-sales price only, or the quantity-sales price only. The rest of this section is devoted to addressing these questions.

We first introduce two critical points in connection with the order-up-to level, which are instrumental in characterizing optimal prices. Let

\[
y_t = \inf \left\{ y : \lambda_t^*(y) > 0 \right\}, \quad \bar{y}_t = \inf \left\{ y : \lambda_t^*(y) > 0, \lambda_t^*(y) - m\lambda_t^*(\lambda_t^*(y)) = 0 \right\}.
\]  

(2.17)

We will identify mild conditions such that $y_t$ is a threshold point below which it is not optimal to sell in the quantity-sales mode, and $\bar{y}_t$ is a threshold point above which the firm will not deploy the unit-sales mode (the threshold points take $+\infty$ if they do not exist). The following lemma specifies the order of these two threshold points and the monotonicity of $\lambda_t^*(y)$.

Lemma 2.1. (a) Under Assumption 2.1, $\lambda_t^*(y)$ increases in the order-up-to level $y$ for any $t = 1, ..., n$.

(b) The two threshold points satisfy $y_t \leq \bar{y}_t$. 

9
In the rest of this section, we show that the two threshold points, together with the optimal base-stock level, determine the optimal selling mode. The following assumption is required as a sufficient condition to identify \( y_t^* \) as the threshold inventory level for the quantity-sales mode.

**Assumption 2.2.** Consider the first partial derivative of \( R_2(\lambda, \lambda_2) \) with respect to \( \lambda_2 \) on the line \( \lambda_2 = 0 \), i.e., \( \frac{\partial R_2}{\partial \lambda_2}(\lambda, 0) \). If there exists some \( \lambda \in (0, m) \) such that \( \frac{\partial R_2}{\partial \lambda_2}(\lambda, 0) > 0 \), then \( \frac{\partial R_2}{\partial \lambda_2}(\lambda', 0) > 0 \) for any \( \lambda' \in (\lambda, m) \).

This assumption stipulates the following. If there exists some overall demand rate \( \lambda \) such that the one-period revenue is strengthened by the introduction of the quantity-sales mode \( (\frac{\partial R_2}{\partial \lambda_2}(\lambda, 0) > 0) \), then the same is also true for any larger overall sales rate \( \lambda' \geq \lambda \). Thus, loosely speaking, if it is better to sell in the quantity-sales mode, then the firm should keep doing so if it wants to sell more. Note that a large family of functions satisfy Assumption 2.2. For example, if \( R_2(\lambda, \lambda_2) \) is supermodular, then \( \frac{\partial R_2}{\partial \lambda_2}(\lambda, \lambda_2) \) is increasing in \( \lambda \) for any given \( \lambda_2 \). Therefore, \( \frac{\partial R_2}{\partial \lambda_2}(\lambda, 0) \) is increasing in \( \lambda \), which immediately yields the condition in Assumption 2.2.

The following assumption is required as a sufficient condition to identify \( y_t \) as the threshold inventory level for the unit-sales mode.

**Assumption 2.3.** Consider the first partial derivative of \( R_1(\lambda_1, \lambda) \) with respect to \( \lambda_1 \) on the line \( \lambda_1 = 0 \), i.e., \( \frac{\partial R_1}{\partial \lambda_1}(0, \lambda) \). If there exists some \( \lambda \in (0, m) \) such that \( \frac{\partial R_1}{\partial \lambda_1}(0, \lambda) \leq 0 \), then \( \frac{\partial R_1}{\partial \lambda_1}(0, \lambda') \leq 0 \) for any \( \lambda' \in (\lambda, m) \).

This assumption states the following. If there exists some overall demand rate \( \lambda \) such that the one-period revenue is weakened by the introduction of the unit-sales mode \( (\frac{\partial R_1}{\partial \lambda_1}(0, \lambda) \leq 0) \), then the same is also true for any larger overall sales rate \( \lambda' \geq \lambda \). Therefore, if it is better to sell only in the quantity-sales mode, then the firm should keep doing so if it wants to sell more.

**Remark 2.3.** Note that a large family of functions satisfies Assumption 2.3. For example, if \( R_1(\lambda_1, \lambda) \) is submodular, then \( \frac{\partial R_1}{\partial \lambda_1}(\lambda_1, \lambda) \) is decreasing in \( \lambda \) for any given \( \lambda_1 \). Therefore, \( \frac{\partial R_1}{\partial \lambda_1}(0, \lambda) \) is decreasing in \( \lambda \), which immediately yields the condition in Assumption 2.3. Technically, that \( R_1(\lambda_1, \lambda) \) is supermodular is almost sufficient for Assumption 2.3 to hold, which is shown in Appendix A (Proposition 0.1).

Although Assumptions 2.1, 2.2, and 2.3 are linked by the equality \( \lambda = \lambda_1 + m \lambda_2 \), we show that they do not imply each other in general. Remarkably, in the next section, it will be shown that these three assumptions can be unified by one assumption under a valuation-based demand model.

With Assumptions 2.2 and 2.3, we have the optimal pricing policy as follows.

**Theorem 2.2. (Pricing Policy)** Consider the optimal order-up-to level \( y_t^* \) (after ordering) in Theorem 2.1, and thresholds \( y_t \) and \( \bar{y}_t \) (defined in (2.17)). If Assumptions 2.1, 2.2, and 2.3 hold, then the optimal selling mode for any period \( t = 1, \ldots, n \), is as follows.

(a) If \( y_t^* \leq y_t \), then the firm sets a uniform price only.

(b) If \( y_t < y_t^* < \bar{y}_t \), then the firm uses both unit-sales and quantity-sales prices.

(c) If \( y_t^* \geq \bar{y}_t \), then the firm uses a quantity-sales price only.

In all cases, the exact optimal prices depend on the optimal inventory level, \( y_t^* \), in period \( t \).
The theorem has already been illustrated in Figure 1. For the reader who is familiar with the inventory-pricing literature, it is easy to perceive that the above analysis can be extended to the infinite-horizon stationary setting, by following a routine procedure. For this reason, we place the analysis in Appendix A.

3 A Valuation-Based Demand Model

The optimal policy derived in the previous section is based on Assumptions 2.1, 2.2, and 2.3. Throughout this section, we assume the valuation-based demand model stipulated in Remark 2.1, and show that Assumptions 2.1, 2.2, and 2.3 can be satisfied by a large class of distributions of V. Moreover, the structural properties we revealed in the previous section are further enhanced and refined to reveal new insights.

3.1 Validation of Assumptions 2.1, 2.2 and 2.3

First, we introduce the concept of the virtual value function (in the case of pure preference uncertainty) developed in the auction and mechanism design literature. If each buyer’s virtual value function is increasing, then we have a regular auction (see Myerson 1981).

**Definition 3.1.** Let V be a random variable with a cumulative distribution function \( G(p) = \Pr(V < p) \) and a probability density function \( g(p) \). Define

\[
 w(p) = p - \frac{1}{h(p)}
\]

where \( h(p) = \frac{g(p)}{1 - G(p)} \) is the failure rate of V. V has an increasing virtual value function if \( w(p) \) is strictly increasing in \( p \).

To understand the virtual value function \( w(p) \) in our setting, note that V denotes utility and \( p \) denotes price, and hence \( x = \Pr(V > p) = 1 - G(p) \) is the demand rate under the unit-sales mode only with a uniform price \( p \). Given \( p = G^{-1}(1 - x) \), we denote \( \zeta(x) = x \times p = xG^{-1}(1 - x) \), which is the revenue function of demand rate \( x \). Then,

\[
 \frac{dy(x)}{dx} = G^{-1}(1 - x) - \frac{\frac{dx}{dG^{-1}(1-x)}}{g(\frac{dx}{dG^{-1}(1-x)})} = G^{-1}(1 - x) - \frac{1}{h(\frac{dx}{dG^{-1}(1-x)})} = p - \frac{1}{h(p)} = w(p).
\]

Therefore, \( w(p) \) measures the sensitivity of revenue to the demand rate. In other words, it represents the revenue increment in response to a unit increase in the demand rate, when units are sold at a uniform price only. Similarly, the failure rate \( h(p) = \frac{g(p)}{1 - G(p)} = \frac{-dx}{dp} \) represents the percentage change in the demand rate in response to a unit change in price \( p \).

Clearly, if \( V \) has an increasing failure rate, i.e., \( h(p) \) is increasing, then \( v \) must have an increasing virtual value function. For readers who are familiar with the generalized failure rate (GFR) defined as \( ph(p) \) (see Lariviere and Porteus 2001 and Lariviere 2006), an increasing virtual value function and an increasing GFR do not imply each other, but both share the class of distributions with an increasing failure rate. It should be noted that under the current context, the GFR \( ph(p) = \frac{-dx}{dp} = \frac{dx}{dp} \) represents the price elasticity of the demand rate.

**Proposition 3.1.** The functions \( R(\lambda_1, \lambda_2) \), \( R_2(\lambda, \lambda_2) = R(\lambda - m\lambda_2, \lambda_2) \), and \( R_1(\lambda_1, \lambda) = R(\lambda_1, \frac{\lambda - \lambda_1}{m}) \) satisfy Assumptions 2.1, 2.2, and 2.3, respectively, if the value of unit product V has an increasing virtual value function. Moreover, if Assumption 2.1 is satisfied, the value of unit product V has an increasing virtual value function.
By this proposition, the three assumptions in Section 2 hold as long as \( V \) has an increasing virtual value function. Because most of the commonly used distributions have an increasing failure rate (see Lariviere 2006), these assumptions are plausible.

**Remark 3.1.** If Assumptions 2.1, 2.2, and 2.3 are violated, i.e., \( V \) does not have an increasing virtual value function, the optimal policy could be significantly different from what is shown in Section 2, and may even be counter-intuitive. This is illustrated by Example 1 in Appendix A.

### 3.2 Optimal Prices

In the market place, we often see promotions such as “buy one get one free”. In this case, the seller uses only the quantity-sales mode because the customer will always take the “free one”. The first part of the ensuing proposition suggests that selling exclusively in the quantity-sales mode is not optimal if the firm’s initial inventory is less than the optimal base-stock level \( S_t \).

**Proposition 3.2.** (a) For each period \( t \), the optimal base-stock level, \( S_t < \overline{y}_t \); i.e., the optimal base-stock level is below the threshold level for the unit sales mode (i.e., above which items will never be sold in the unit-sales mode).

(b) \( p_{1t}^* \) is decreasing in \( y_t^* < \overline{y}_t \) and \( p_{2t}^* \) is decreasing in \( y_t^* > \overline{y}_t \). Moreover, \( p_{2t}^* - p_{1t}^* \) always decreases with the optimal order-up-to level \( y_t^* \).

An alternative interpretation of Part (a) is as follows. For any given period and initial inventory level, if it is optimal to make an inventory replenishment, then it would never be optimal to sell exclusively in the quantity-sales mode. It should also be noted that the firm may still sell exclusively in the quantity-sales mode if \( y_t^* > \overline{y}_t \). However, in the stationary setting, such a case can be ruled out in the steady state because the base-stock level will be a constant. The second part of the proposition is intuitive: the higher the inventory level, the lower the selling price of the corresponding sales mode, and the optimal quantity discount price \( p_{2t}^* \) should decrease faster than the optimal unit-sales price \( p_{1t}^* \) when the inventory increases.

Let \( \Delta p_t^* = m p_{1t}^* - p_{2t}^* \). If \( \Delta p_t^* > 0 \), then a quantity discount exists, whereas if \( \Delta p_t^* < 0 \), the quantity surcharge is applied. (Although consumers expect a larger quantity to be priced in a quantity-discount fashion, quantity surcharges, which exist when the unit price is higher for a large-size package than for a small one, are common in the retail grocery market. Here, \( m \) can be a non-integer number. See Sprott et al. (2003) for empirical evidence and the references therein.)

**Proposition 3.3.** (Condition of Quantity Discount) If \( V \) has an increasing failure rate, then \( \Delta p_t^* \geq 0 \) for any \( y_t^* \in (\overline{y}_t, 2 \overline{y}_t) \).

This proposition asserts a condition under which quantity discounts must be applied. (Although it is intuitive, the result is nontrivial: if \( V \) does not have an increasing failure rate, then we may have \( \Delta p_t^* < 0 \), i.e., a quantity surcharge can be optimal; see Example 2 in Appendix A.) To explain the insight behind this proposition, we need to understand the meaning of failure rate \( h(p) \) in the current context. As explained
earlier, \(x = P(V > p) = 1 - G(p)\) represents the demand rate under price \(p\), and the failure rate \(h(p) = g(p)/(1 - G(p)) = \frac{-dx}{dp}\) represents the percentage change in the demand rate in response to a unit change in price \(p\). If \(h(p)\) increases with \(p\), the percentage change is an increasing function of the unit-sales price \(p\), which implies that demand decreases at a faster rate when we increase price \(p\). Therefore, the firm should use quantity discounts.

As defined earlier, the value \(\triangle p_t^* (\geq 0)\) represents the amount of the quantity discount. Intuitively, the higher the inventory level, the more quantity discount the firm should offer. Surprisingly, this is not always true.

**Proposition 3.4. (Amount of Quantity Discounts)** Suppose that \(V\) has an increasing failure rate. For any \(y_t^* \in (y_t, \bar{y}_t)\), \(\triangle p_t^*\) is increasing in \(y_t^*\) if \(w(p)\) is concave, and \(\triangle p_t^*\) is decreasing in \(y_t^*\) if \(w(p)\) is convex.

For any \(y_t^* \in (y_t, \bar{y}_t)\), the optimal selling strategy is to use both selling modes (see Theorem 2.2). Intuition tells us that the quantity discount \(\triangle p_t^*\) should increase in \(y_t^*\). However, Proposition 3.4 asserts a definite condition under which reverse monotonicity exists when \(w(p)\) is convex. As the virtual value function \(w(p)\) measures the sensitivity of revenue to the demand rate, Proposition 3.4 may be interpreted as follows: if the sensitivity of revenue to the demand rate has a decreasing (increasing) growth rate with respect to price, then the amount of the quantity discount increases (decreases) with the inventory level.

Next, we demonstrate Proposition 3.4 by considering several different distributions of \(V\) (all of which have an increasing failure rate). If \(V\) follows a uniform distribution or exponential distribution, we can show that \(w(p)\) is a linear function of \(p\), which is both convex and concave. Therefore, Proposition 3.4 implies that \(\triangle p_t^*\) is a constant that is independent of the inventory level. More specifically, \(\triangle p_t^* = \frac{m-a}{2}\) for the uniform distribution and \(\triangle p_t^* = (m-a)u\) for the exponential distribution with expectation \(u\). If \(V\) follows a triangular distribution, then it can be shown that \(w(p)\) is concave and hence the amount of quantity discount \(\triangle p_t^*\) increases with the inventory level. Suppose that \(V\) follows a truncated logistic distribution with a support \([0, \overline{v}]\), i.e., \(G(x) = (\tilde{G}(x) - \tilde{G}(0))/(\tilde{G}(\overline{v}) - \tilde{G}(0))\), where \(\tilde{G}(x) = \frac{1}{1+e^{-(x-\mu)/s}}\), which is the cumulative distribution function of the logistic distribution with location \(\mu\) and scale \(s\). We can show that if \(\tilde{G}(\overline{v}) \leq 1/2\), then \(w(p)\) is concave and the amount of quantity discount \(\triangle p_t^*\) increases with the inventory. However, if \(\frac{1}{2} \leq \tilde{G}(0) \leq \tilde{G}(\overline{v}) \leq \frac{\tilde{G}(0)^2}{1-2\tilde{G}(0)+2\tilde{G}(0)^2}\), then \(w(p)\) is convex and hence the amount of quantity discount \(\triangle p_t^*\) should decrease with the inventory level.

### 4 Value of Quantity-based Price Differentiation

The purpose of this section is to study the additional profit to be gained from offering two different prices compared with charging a uniform price, which we refer to as the value of quantity-based price differentiation, or the value of price differentiation in short. We first explore how the value of price differentiation depends on the system parameters and under what conditions this value is significant.

For ease of exposition, we focus on the infinite horizon model with stationary inputs. Moreover, we assume that the demand function is derived from the utility model introduced in the previous section, and the value of
one unit of product $V$ is assumed to have an increasing virtual value function. In this setting, all the structural properties developed earlier hold in a stationary fashion. We also assume that the inventory holding/shortage cost for any period $t$ takes the form $\ell(y) = hy^+ + sy^-$, where $h$ and $s$ denote the unit inventory holding and shortage cost, respectively. To avoid triviality, we impose $s > (1-\theta)c$ (where $\theta$ is the discount factor).

Without the quantity-sales mode, the firm applies uniform pricing. Then, by setting $\lambda_2$ to be zero, the one-period expected revenue for the model without quantity sales becomes
\[
\hat{\Gamma}(\lambda) = \lambda G^{-1}(1 - \lambda) \cdot \mu.
\] (4.1)

Note that the one-period revenue function $\hat{\Gamma}(\lambda)$ can also be obtained by a utility model similar to that introduced in Section 3. Let $p$ denote the unit price quoted by the firm. Following from the previous section, the demand rate $\lambda(p) = \Pr(V > p) = 1 - G(p)$ for any given $p$, or equivalently, $p(\lambda) = G^{-1}(1 - \lambda)$ for any $\lambda \in [0,1]$.

It follows that the one-period expected revenue is $\hat{\Gamma}(\lambda)$ in (4.1). Proposition 3.1 implies that $\hat{\Gamma}(\lambda)$ is strictly concave in $\lambda \in [0,1]$. The dynamic program for the uniform pricing problem can easily be formulated by modifying the counterpart for the two selling modes, which is omitted here.

Suppose that the firm starts with zero inventory. Let $V(0)$ (and $\hat{V}(0)$) be the optimal expected profits for the model with two selling modes (and with the unit-sales mode only). Denote the value of price differentiation by $\Delta = V(0) - \hat{V}(0)$.

**Lemma 4.1.** The value of price differentiation $\Delta$ equals
\[
\Delta = \frac{1}{1 - \theta} \left[ J_\infty(S, \lambda^o) - \hat{J}_\infty(\tilde{S}, \tilde{\lambda}^o) \right],
\]
where
\[
J_\infty(y, \lambda) = (\theta - 1)c y - E[\ell(y - \lambda D - \epsilon)] + \Gamma(\lambda) - \theta c \mu \lambda,
\]
\[
\hat{J}_\infty(y, \lambda) = (\theta - 1)c y - E[\ell(y - \lambda D - \epsilon)] + \hat{\Gamma}(\lambda) - \theta c \mu \lambda,
\]
\[
(S, \lambda^o) = \arg \max_{y, \lambda \in [0,1]} \left\{ J_\infty(y, \lambda) \right\} \quad \text{and} \quad (\tilde{S}, \tilde{\lambda}^o) = \arg \max_{y, \lambda \in [0,1]} \left\{ \hat{J}_\infty(y, \lambda) \right\}.
\] (4.2)

Intuitively, when considering both the unit-sales and quantity-sales modes, the firm should choose a higher base-stock level ($S$) and a higher expected demand ($\lambda^o$) than when selling exclusively in the unit-sales mode ($\tilde{S}$ and $\tilde{\lambda}^o$ respectively). This is confirmed by the next proposition.

**Proposition 4.1.** $S \geq \tilde{S}$ and $\lambda^o \geq \tilde{\lambda}^o$. As a result, the expected demands $\lambda^o \mu \geq \tilde{\lambda}^o \mu$.

To understand how quantity discounts can enhance revenue, we consider the case in which the unit value of product $V$ follows a uniform distribution in $[0,1]$, $m = 2$, $a = 1.5$, and $\mu = 1$. Suppose that the optimal demand rate is $\tilde{\lambda} = \frac{3}{4}$ if we use the unit-sales-only mode. Then the respective price $p_1(\tilde{\lambda}) = (1 - \tilde{\lambda}) = 1/4$ and the revenue is $\frac{1}{16}$. For ease of comparison, we keep the demand rate unchanged at $\tilde{\lambda} = \frac{3}{4}$ but quote both the unit-sales and quantity-sales prices $p_1$ and $p_2$. These two prices can be represented as functions of the unit-sales market share $\lambda_1$. Equation (2.6) implies that $p_1(\lambda_1, \frac{3-\lambda_1}{m}) = \frac{5}{8} - \frac{\lambda_1}{2}$ and $p_2(\lambda_1, \frac{3-\lambda_1}{m})/m = \frac{15}{32} - \frac{\lambda_1}{8}$, which correspond to the bold black and gray lines in Figure 2, respectively. To maximize the revenue (with
Figure 2: Additional Revenue Obtained by Quantity-based Price Differentiation

the overall demand rate is fixed at \( \hat{\lambda} = \frac{3}{4} \), we can show that the optimal market share for the unit-sales 
\( \lambda^*_1 \left( \frac{3}{4} \right) = \frac{1}{72} \) with the respective price \( p_1(\lambda^*_1(\frac{3}{4})) = \frac{7}{12} \), the quantity-sales price \( p_2(\lambda^*_1(\frac{3}{4})), \frac{4/3 - \lambda^*_1(\frac{3}{4})}{m} / m = \frac{11}{24} \) and its share \( \left( \frac{3}{4} - \lambda^*_1(\frac{3}{4}) \right) = \frac{2}{3} \). As a result, the expected revenue is \( \frac{17}{48} \), which almost doubles the revenue of \( \frac{3}{16} \) generated by the unit-sales-only mode. Here, \( p_1() \times \lambda^*_1() \) is represented by two rectangles A and C, and 
\( p_2() \times (\frac{3}{4} - \lambda^*_1(\frac{3}{4})) \) is represented by the two rectangles B and D. Therefore, if prices \( p_1 \) and \( p_2 \) are both quoted, 
the expected revenue is the total area of rectangles A, B, C, and D. Obviously, we obtain additional revenue 
of \( \frac{1}{6} \), i.e., the area of rectangles A and B, by adopting quantity-based price differentiation.

4.1 Comparative Statics

Lemma 4.1 and Proposition 4.1 help us establish the following monotonic property of the value of price differ-
teniation \( \Delta \) with respect to the unit order cost \( c \).

Proposition 4.2. The value of quantity-based price differentiation, \( \Delta \), decreases in per-unit ordering cost \( c \).

When the unit ordering cost \( c \) increases, the cost will increase in both systems. However, the cost increments 
in both systems depend on the base-stock levels \( S \) and \( \hat{S} \). Proposition 4.1 implies that \( S \geq \hat{S} \) and hence the 
value of price differentiation decreases with the unit ordering cost \( c \).

Similarly, when the unit inventory holding cost \( h \) increases, the cost increments in both systems depend on 
their expected leftover inventory. For both the multiplicative demand \( (\varepsilon_t \equiv 0) \) and the additive model \( (D_t \) is deterministic), we can show that the expected leftover inventory is higher in the system with quantity-based 
price differentiation. Hence, the value of price differentiation decreases with the unit inventory holding cost \( h \).
For the general demand model, we can find examples in which the expected leftover inventory in the system 
with quantity-based price differentiation is not always higher and hence this monotonicity does not always hold.
The same results apply to the unit shortage cost \( s \). When \( s \) increases, the cost increments in both systems
depend on their expected shortages. For both the multiplicative demand and the additive model, we can show that the expected shortages is higher in the system with quantity-based price differentiation. Therefore, the value of price differentiation decreases with the unit shortage cost $s$. However, we can find examples in which this monotonicity does not always hold for the general demand model.

Another important parameter in the model with quantity-discount sales is $a$, which determines the marginal utility rate for $m$ units. Intuitively, the larger the value of $a$, the higher the value of price differentiation. The following proposition confirms this intuition.

Proposition 4.3. $\Delta$ increases in the valuation parameter $a$.

4.2 Numerical Study

The primary objective of this numerical study is to quantify the benefit of deploying quantity-based pricing and to identify the conditions under which quantity-based pricing (or simply price differentiation) can bring significant additional value compared to uniform pricing. This is done mainly under the infinite-horizon setting. We also investigate the finite-horizon setting to see how the length of the planning horizon affects the value of price differentiation, and to compare the value of price differentiation with the value of shifting from the fixed pricing strategy to the dynamic pricing strategy.

We assume the random utility $V$ follows a class of distributions $G(p) = \Pr(V < p) = p^k$, where $0 < k \leq 1$ with support $[0,1]$. Note that for any $p \in [0,1]$, $p^{k_1} \leq p^{k_2}$ for $k_1 \leq k_2$, which implies that the utility $v$ stochastically decreases with $k$. (That is, the higher the value of $k$, the lower the value of the product per unit.) In the numerical study, we fix $\theta = 0.98$, $m = 2$, $D \sim U[0,200]$, and $\epsilon \sim U[-10,10]$. The baseline setting is $a = 1.6$, $c = 0.4$, $h = 0.008$, and $s = 0.03$ (linear holding and shortage costs per unit per period). The value of price differentiation is measured by $\frac{\Delta b}{b(0)} \times 100\%$.

In Table 2, we present the value of price differentiation with different parameter values for $a$, $c$, $h$, and $s$. The first column shows different $k$ values, and the rest for the respective parameters. First, a common observation from Table 2 is that the overall profit improvement is rather significant, ranging from 0 to 63%. Low profit improvement only occurs when the marginal utility increment (i.e., $a - 1$) from the quantity purchase is low (i.e., the second unit brings only 40-50\% more utility) or when the ordering cost is large (i.e., $c \geq 0.5$). From Table 2, we find that the value of price differentiation decreases with $k$. This is easy to understand because the per-unit utility of the product stochastically decreases with $k$.

Another observation is that the value of price differentiation is very sensitive to $a$ and $c$, whereas it is less sensitive to the unit-holding cost $h$ and the unit-penalty cost $s$. As $aV$ is the utility for $m$ units, we can expect the quantity-sales mode to bring a significant profit increase when $a$ is large. When the unit ordering cost $c$ increases, the cost increases faster in the system with price differentiation because of the higher base-stock level. As this is a direct effect on cost, the value of price differentiation diminishes when $c$ increases. Certainly, a change in the inventory holding/shortage cost affects the value of price differentiation through its effect on the optimal order-up-to level. However, the firm makes inventory and pricing decisions simultaneously, which ensures a better match between supply and demand. That is why the effect of $h$ and $s$ is much smaller than $a$. 

16
and $c$.

Table 2: Effects of Parameters on the Value of Price Differentiation

<table>
<thead>
<tr>
<th>$k$</th>
<th>Percentage of Improvement %</th>
<th>$k$</th>
<th>Percentage of Improvement %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Utility Parameter $a$</td>
<td></td>
<td>Per-unit Purchasing Cost $c$</td>
</tr>
<tr>
<td>0.05</td>
<td>1.4 1.5 1.6 1.7 1.8</td>
<td>0.05</td>
<td>0.1 0.2 0.3 0.5 0.6</td>
</tr>
<tr>
<td>0.1</td>
<td>0 9.9 26.7 44.6 62.9</td>
<td>0.1</td>
<td>53.8 46.3 36.7 9.2 0</td>
</tr>
<tr>
<td>0.4</td>
<td>0 5.5 19.2 37.0 56.8</td>
<td>0.4</td>
<td>52.1 42.9 31.9 6.5 0</td>
</tr>
<tr>
<td>0.8</td>
<td>0 4.9 17.8 35.1 55.2</td>
<td>0.8</td>
<td>51.1 41.3 30.1 5.9 0</td>
</tr>
<tr>
<td>1</td>
<td>0 4.8 17.4 34.7 54.8</td>
<td>1</td>
<td>50.8 40.9 29.6 5.7 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Per-unit Holding Cost $h$</th>
<th></th>
<th>Per-unit Shortage Cost $s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.002 0.004 0.006 0.010 0.012</td>
<td>0.05</td>
<td>0.01 0.02 0.04 0.05 0.06</td>
</tr>
<tr>
<td>0.1</td>
<td>27.1 26.9 26.8 26.6 26.5</td>
<td>0.1</td>
<td>25.2 24.5 24.0 23.8 23.7</td>
</tr>
<tr>
<td>0.4</td>
<td>19.6 19.4 19.3 19.2 19.1</td>
<td>0.4</td>
<td>20.2 19.6 19.1 19.0 18.9</td>
</tr>
<tr>
<td>0.8</td>
<td>18.1 17.9 17.9 17.7 17.6</td>
<td>0.8</td>
<td>18.7 18.1 17.6 17.5 17.4</td>
</tr>
<tr>
<td>1</td>
<td>17.7 17.6 17.5 17.4 17.3</td>
<td>1</td>
<td>18.3 17.7 17.3 17.2 17.1</td>
</tr>
</tbody>
</table>

It is well known that the benefit of dynamic pricing (i.e., uniform pricing) quickly vanishes as the planning horizon extends (see Federgruen and Heching 1999) because the optimal base-stock level and posted price converge to a constant. Will this pattern remain true for the current setting? To address this issue, we turn to a finite-horizon with $N$ periods (see Table 3).

The value of price differentiation is measured as $\frac{V_1(0) - \hat{V}_1(0)}{\hat{V}_1(0)} \times 100\%$. We denote the maximal profit generated by static pricing by $\tilde{V}_1(0)$. Hence, $\frac{(V_1(0) - \hat{V}_1(0))}{\tilde{V}_1(0)} \times 100\%$ measures the benefit of dynamic pricing compared with static pricing. To see how the results vary with the parameter values, 100 instances were randomly drawn from the following ranges: $a \sim U(1.4, 1.8)$, $c \sim U(0.1, 0.6)$, $h \sim (0.002, 0.012)$, and $s \sim (0.01, 0.06)$; the demand distribution (of $D_t$) is $U(100 - \delta, 100 + \delta)$, where $\delta$ is uniformly generated in $\{5, 6, ..., 100\}$. The salvage value at the end of the planning horizon is $\tilde{h}c$, where $\tilde{h} \sim U(0, 0.6)$, and the penalty cost for shortage at the end of the planning horizon is set to $c$. Also note that all the parameters are generated independently. Finally, $m = 2$ and $\varepsilon \sim U[-10, 10]$. Table 3 summarizes the results for the average and the maximum of the benefits among the 100 randomly generated instances for different values of $N$ and $k$.

Table 3: Percentage of Improvement by Different Pricing Strategies: Average and Maxima

<table>
<thead>
<tr>
<th>$N \backslash k$</th>
<th>Dynamic Uniform Pricing vs. Static Pricing</th>
<th>Quantity-based Pricing vs. Dynamic Uniform Pricing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.05</td>
<td>0.1</td>
</tr>
<tr>
<td>1</td>
<td>(0,0)</td>
<td>(0,0)</td>
</tr>
<tr>
<td>2</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>3</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>4</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>5</td>
<td>(0.01)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>6</td>
<td>(0.01)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>7</td>
<td>(0.01)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>8</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>9</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>10</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
</tbody>
</table>
Federgruen and Heching (1999)’s finding that the value of dynamic pricing vanishes as the planning horizon extends is also confirmed by the left panel of Table 3. In contrast, the right panel of Table 3 shows that firms adopting price differentiation may enjoy long-term benefits, i.e., the value of price differentiation increases with the total number of periods in the planning horizon. In this case, the price differentiation helps to push up demand in each period. When the number of periods increases, the total demand pushed up by the price differentiation during the planning horizon also increases, and hence the firm receives more benefit from price differentiation.

5 Conclusion and Extensions

This paper proposes an inventory-pricing model, in which the optimal simultaneous decisions on inventory control and selling/pricing are based on a simple structure. In particular, the following results are of managerial relevance. First, quantity-based price differentiation can produce a large profit improvement if the firm shifts from dynamic pricing with the unit-sales mode only. This benefit is significant when the ordering, holding, and shortage costs are relatively small, or when the marginal rate of utility is relatively large with respect to the purchased quantity. Second, the quantity-sales mode alone cannot be optimal when inventory replenishment is required. Third, the amount of quantity discount does not have to increase with the inventory level because the direction of monotonicity depends on the sensitivity of revenue to the demand rate. A research takeaway is a set of assumptions for the demand models that allow us to solve complex, dynamic pricing, and multi-selling mode problems. An equally important technical takeaway is the concept of the virtual value function, which has its roots in the auction and mechanism design literature and is applicable to inventory-pricing problems.

One possible future research direction is to consider a model without a replenishment opportunity, i.e., the revenue management version of our model. Monahan et al. (2004) studied the problem of (dynamic) uniform pricing and obtained interesting structural results. It would be interesting to include quantity-based price differentiation in their model.

References

Lemma 5.1. If $f(\cdot)$ is a concave function, then $f(au - bv)$ is supermodular and jointly concave in $(u, v)$ for any fixed nonnegative numbers $a$ and $b$.

Proof. For any $u_1 < u_2$ and $v_1 < v_2$, we have $au_2 - bv_1 \geq au_2 - bv_2$ and $au_1 - bv_1 \geq au_1 - bv_2$ since $a, b \geq 0$. Also note that $(au_2 - bv_1) - (au_1 - bv_1) = (au_2 - bv_2) - (au_1 - bv_2) = a(u_2 - u_1) > 0$. The concavity of $f(\cdot)$ shows that $f(au_2 - bv_1) - f(au_1 - bv_1) \leq f(au_2 - bv_2) - f(au_1 - bv_2)$. Therefore, $f(au - bv)$ is supermodular in $(u, v)$. The joint concavity of $f(au - bv)$ can be directly obtained from Lemma 1 in Gallego and Hu (2004).

Proof of Theorem 1. This result can be proved by induction. We consider the induction assumption that $W_{t+1}(x)$ is concave, which is true for $t = n$, as $W_{n+1}(x) = V_{n+1}(x) - cx$, where $V_{n+1}(x)$ is assumed to be concave.

Consider the function $J_t(y, \lambda)$ defined in (14). Since $W_{t+1}(x)$ is concave and $\ell(y)$ is convex, Lemma 5.1 implies that $-E[\ell(y - \lambda D_t - \varepsilon_t)] + \theta E[V_{t+1}(y - \lambda D_t - \varepsilon_t)]$ is jointly concave in $(y, \lambda)$. As $\Gamma(\lambda)$ is concave and the rest terms in $J_t(y, \lambda)$ are linear, $J_t(y, \lambda)$ is jointly concave in $(y, \lambda)$.

Closely following the proof of Proposition B-4 in Heyman and Sobel (1984), we can show that $J_t(y, \lambda^*_t(y)) = \max_{0 \leq \lambda \leq m} \{J_t(y, \lambda)\}$ is a concave function of $y$. As $y^*_t = \arg\max_{y \geq 0} \{J_t(y, \lambda^*_t(y))\}$, the concavity of $J_t(y, \lambda^*_t(y))$ in $y$ yields that $y^*_t = S_t$ for any $x \leq S_t$ and $y^*_t = x$ for any $x > S_t$.

Similar to Proposition B-4 in Heyman and Sobel (1984), we can establish that $W_t(x)$ is concave in $x$. □

Proof of Lemma 1. (a) As $W_t(x)$ is a concave function of $x$ for any $t = 1, ..., n$ (see the proof of Theorem 1) and $\ell(y)$ is a convex function of $y$, the definition of $J_t(y, \lambda)$ and Lemma 5.1 imply that $J_t(y, \lambda)$ is supermodular in $(y, \lambda) \in (-\infty, +\infty) \times [0, m]$. Then, the desired result follows.

(b) To see $y^*_t \leq \bar{y}_t$, we can write $\bar{y}_t = \inf\{y : \lambda^*_2(\lambda^*_t(y)) > 0, \lambda^*_t(y) - m\lambda^*_2(\lambda^*_t(y)) = 0\}$. The definition of $\bar{y}_t$ has one more constraint than the definition of $y^*_t$, which implies that $y^*_t \leq \bar{y}_t$ since we are taking “inf”. □

The following lemma follows immediately from the definition of $\mathcal{B}$, which will be useful soon.

Lemma 5.2. The set $\mathcal{B}$ defined in (10) can be rewritten as

\[ \mathcal{B} = \left\{ (\lambda, \lambda_2) : 0 \leq \lambda \leq m, \Delta_2(\lambda) \leq \lambda_2 \leq \frac{\lambda}{m} \right\}. \]
where $\Delta_2(\lambda)$ is a nonnegative, convex and increasing function with $\Delta_2(\lambda) = 0$ for $0 \leq \lambda \leq 1$ and $\Delta_2(\lambda) = \frac{1}{m-1} \lambda - \frac{1}{m-1}$ for $1 < \lambda \leq m$.

The next proposition provides an alternative to Assumption 3.

**Proposition 5.1.** Under Assumptions 1 and 2, if $R_1(\lambda_1, \lambda)$ is supermodular in the set $[0, 1] \times [0, m]$ and $\Delta > 0$, then Assumption 3 is satisfied, where

$$\lambda = \begin{cases} \inf \left\{ \lambda \in (0, 1) : \frac{\partial R_2}{\partial \lambda_2}(\lambda, 0) > 0 \right\} & \text{if } \left\{ \lambda \in (0, 1), \frac{\partial R_2}{\partial \lambda_2}(\lambda, 0) > 0 \right\} \neq \emptyset, \\ 1 & \text{if } \left\{ \lambda \in (0, 1), \frac{\partial R_2}{\partial \lambda_2}(\lambda, 0) > 0 \right\} = \emptyset. \end{cases} \quad (5.1)$$

**Proof.** Since the domain of the function $R(\lambda_1, \lambda_2)$ is $\mathcal{U}$, the domain of function $R_1(\lambda_1, \lambda)$ can be defined as $\mathcal{U} = \{(\lambda_1, \lambda) : (\lambda_1, \lambda_2) \in \mathcal{U}\}$. Similar to the definition of $\lambda^*_2(\lambda)$ in (12), given any $\lambda \in [0, m]$, the optimal market share of the unit-sales price is $\lambda^*_1(\lambda)$ such that

$$\lambda^*_1(\lambda) = \arg\max_{\lambda_1 \in \mathcal{U}(\lambda)} \left\{ R_1(\lambda_1, \lambda) \right\}, \text{ where } \mathcal{U}(\lambda) = \left\{ \lambda_1 : (\lambda_1, \lambda) \in \mathcal{U} \right\}.$$

First, we prove that $\lambda^*_1(\lambda) + m\lambda^*_2(\lambda) = \lambda$ for all $\lambda \in [0, m]$. Suppose for contradiction that $\lambda^*_1(\lambda) + m\lambda^*_2(\lambda) \neq \lambda$ for some $\lambda \in [0, m]$. The definition of $\lambda^*_1(\lambda)$ implies that $(\lambda^*_1(\lambda), \lambda) \in \mathcal{U}$. Applying the definitions of $\mathcal{B}$ and $\mathcal{U}$, we obtain $(\lambda^*_1(\lambda), \frac{\lambda - \lambda^*_1(\lambda)}{m}) \in \mathcal{B}$, and hence $(\lambda, \frac{\lambda - \lambda^*_1(\lambda)}{m}) \in \mathcal{B}$.

Since $\lambda^*_2(\lambda)$ is the maximizer of $R_2(\lambda, \lambda_2)$ for given $\lambda$, we know that $R_2(\lambda, \lambda^*_2(\lambda)) > R_2\left(\lambda, \frac{\lambda - \lambda^*_1(\lambda)}{m}\right)$, where the strict inequality follows from the strict concavity of $R_2(\lambda, \lambda_2)$. The definition of $R_2(\lambda, \lambda_2)$ yields that

$$R(\lambda - m\lambda^*_2(\lambda), \lambda^*_2(\lambda)) = R_2(\lambda, \lambda^*_2(\lambda)) \quad \text{and} \quad R\left(\lambda^*_1(\lambda), \frac{\lambda - \lambda^*_1(\lambda)}{m}\right) = R_2\left(\lambda, \frac{\lambda - \lambda^*_1(\lambda)}{m}\right),$$

and hence we obtain $R(\lambda - m\lambda^*_2(\lambda), \lambda^*_2(\lambda)) > R\left(\lambda^*_1(\lambda), \frac{\lambda - \lambda^*_1(\lambda)}{m}\right)$. Symmetrically, applying the definition of $\lambda^*_1(\lambda)$, we can also show that $R\left(\lambda^*_1(\lambda), \frac{\lambda - \lambda^*_1(\lambda)}{m}\right) > R(\lambda - m\lambda^*_2(\lambda), \lambda^*_2(\lambda))$, which implies a contradiction. Hence, $\lambda^*_1(\lambda) + m\lambda^*_2(\lambda) = \lambda$ for all $\lambda \in [0, m]$.

The definition of $\lambda$ in (5.1) implies that $\lambda^*_2(\lambda) = 0$ for any $\lambda \in (0, \Delta)$. Therefore, we obtain $\lambda^*_1(\lambda) = \lambda > 0$ for any $\lambda \in (0, \Delta)$. Next, we will show that $\lambda^*_1(\lambda) > 0$ for any $\lambda \in (0, m)$ is equivalent to $\frac{\partial R_1}{\partial \lambda_1}(0, \lambda) > 0$.

According to the definitions of $\mathcal{B}$ and $\mathcal{U}$, Lemma 5.2 implies that

$$\mathcal{U} = \left\{ (\lambda_1, \lambda) : 0 \leq \lambda \leq m, \ 0 \leq \lambda_1 \leq \lambda - m\Delta_2(\lambda) \right\}. \quad (5.2)$$

As a result, $\mathcal{U}(\lambda)$ is equal to the interval $[0, \lambda - m\Delta_2(\lambda)]$. Lemma 5.2 also shows that $\lambda - m\Delta_2(\lambda) = 0$ when $\lambda = 0$ or $m$. Consider $\lambda_{1,\text{max}}$ defined as

$$\lambda_{1,\text{max}} = \max \left\{ \lambda_1 : (\lambda_1, \lambda_2) \in \mathcal{U} \right\}. \quad (5.3)$$

Note that the firm needs to consider whether to quote the unit-sales price, which implies that the market share for the unit-sales price, i.e., $\lambda_1$, can be strictly positive in the feasible region $\mathcal{U}$. It follows that there exists some $(\lambda_1, \lambda) \in \mathcal{U}$ such that $\lambda_1 > 0$, and hence we have $\lambda_{1,\text{max}} > 0$. 

21
As \((0,0) \in \mathcal{B}\) and \((m,1) \in \mathcal{B}\), we obtain \((0,0) \in \mathcal{C}\) and \((0,m) \in \mathcal{C}\) by the definition of \(\mathcal{C}\). The definition of \(\lambda_{1,\max}\) in (5.3) also means that there exists some \(\hat{\lambda} \in (0,m)\) such that \((\lambda_{1,\max}, \hat{\lambda}) \in \mathcal{C}\). Since \(\mathcal{C}\) is a linear transformation of the convex set \(\mathcal{A}\), it is a convex set as well. Therefore, the triangle defined by three points \((0,0)\), \((0,m)\) and \((\lambda_{1,\max}, \hat{\lambda})\) is a subset of \(\mathcal{C}\), and hence

\[
\frac{\lambda_{1,\max}}{\lambda} \lambda \leq \lambda - m\Lambda_2(\lambda) \text{ if } \lambda \in (0, \hat{\lambda}], \quad \frac{\lambda_{1,\max}}{m - \lambda} (m - \lambda) \leq \lambda - m\Lambda_2(\lambda) \text{ if } \lambda \in (\hat{\lambda}, m],
\]

which essentially states that the curve \(\lambda - m\Lambda_2(\lambda)\) is above the line segment connecting \((0,0)\) and \((\lambda_{1,\max}, \hat{\lambda})\) and the line segment connecting \((\lambda_{1,\max}, \hat{\lambda})\) and \((0,m)\). Recall that \(\lambda_{1,\max} > 0\). It follows that \(\lambda - m\Lambda_2(\lambda) > 0\) for any \(\lambda \in (0,m)\). The properties about the set \(\mathcal{C}(\lambda)\) indicates that

\[
\lambda_1^*(\lambda) = \arg\max_{0 \leq \lambda_1 \leq m - \lambda m_2(\lambda)} \{R_1(\lambda_1, \lambda)\}, \quad \text{where } \lambda - m\Lambda_2(\lambda) > 0 \text{ for any } \lambda \in (0,m).
\]

(5.4)

Note that \(R_1(\lambda_1, \lambda)\) is a concave function as it is a linear transformation of the concave function \(R(\lambda_1, \lambda_2)\). Therefore, (5.4) implies that for any \(\lambda \in (0,m),

\[
\lambda_1^*(\lambda) > 0 \text{ for any } \frac{\partial R_1}{\partial \lambda_1}(0, \lambda) > 0, \quad \lambda_1^*(\lambda) = 0 \text{ for any } \frac{\partial R_1}{\partial \lambda_1}(0, \lambda) \leq 0.
\]

(5.5)

Note that \(\lambda_1^*(\lambda) = \lambda > 0\) for any \(\lambda \in (0, \Lambda]\). Also note that \(\lambda \leq m\) by definition. According to (5.5), for any \(\lambda \in (0, \Lambda]\), \(\lambda_1^*(\lambda) > 0\) implies that \(\frac{\partial R_1}{\partial \lambda_1}(0, \lambda) > 0\).

When \(R_1(\lambda_1, \lambda)\) is supermodular in \([0,1] \times [0,m]\), \(\frac{\partial R_1}{\partial \lambda_1}(0, \lambda)\) is increasing in \(\lambda \in [0,m]\) for any given \(\lambda_1 \in [0,1]\). Therefore, the positivity of \(\frac{\partial R_1}{\partial \lambda_1}(0, \lambda)\) in \([0, \Lambda]\) yields that \(\frac{\partial R_2}{\partial \lambda_2}(0, \lambda) > 0\) for any \(\lambda \in [\Lambda, m]\). As a result, \(\frac{\partial R_2}{\partial \lambda_2}(0, \lambda) > 0\) for any \(\lambda \in (0,m)\), which is a special case of Assumption 3.

**Proof of Theorem 2.** We first prove that \(y_{\hat{\lambda}}\) is a threshold point for the quantity-sales mode, i.e., \(\lambda_2^*(\lambda)(y) = 0\) for \(y \leq y_{\hat{\lambda}}\), and \(\lambda_2^*(\lambda)(y) > 0\) for \(y > y_{\hat{\lambda}}\). We want to show that \(\lambda_2^*(\lambda) = 0\) for any \(\lambda \in [0, \Lambda]\), and \(\lambda_2^*(\lambda) > 0\) for any \(\lambda \in (\Lambda, m]\). Recall the function \(\lambda_2^*(\lambda)\) defined in (12). As \(\lambda = \lambda_1 + m\lambda_2\) and \(\lambda_1 \in [0,1]\), we have \(\lambda_2^*(\lambda) > 0\) for \(\lambda \in (1, m]\). Next, we just prove \(\lambda_2^*(\lambda) = 0\) for any \(\lambda \in [0, \Lambda]\) and \(\lambda_2^*(\lambda) > 0\) for any \(\lambda \in (\Lambda, 1]\).

Note that \(R_2(\lambda, \lambda_2)\) is strictly concave in \(\lambda_2\) for any \(\lambda \in (0,1)\). Therefore, for any \(\lambda \in (0,1),\) we have

\[
\lambda_2^*(\lambda) = 0 \text{ for any } \frac{\partial R_2}{\partial \lambda_2}(\lambda, 0) \leq 0, \quad \lambda_2^*(\lambda) > 0 \text{ for any } \frac{\partial R_2}{\partial \lambda_2}(\lambda, 0) > 0.
\]

Assumption 2 yields that \(\frac{\partial R_2}{\partial \lambda_2}(\lambda, 0) \leq 0\) for any \(\lambda \in (0, \Lambda]\), and \(\frac{\partial R_2}{\partial \lambda_2}(\lambda, 0) > 0\) for any \(\lambda \in (\Lambda, 1]\); i.e., \(\lambda_2^*(\lambda) = 0\) for any \(\lambda \in (0, \Lambda]\), and \(\lambda_2^*(\lambda) > 0\) for any \(\lambda \in (\Lambda, 1]\). Also note that \(\mathcal{B}(0) = \{0\}\), and hence \(\lambda_2^*(0) = 0\), where \(\mathcal{B}(\lambda)\) is defined in (11). Since \(\lambda_2^*(y)\) is increasing in \(y\), the definition of \(y_{\hat{\lambda}}\) in (17) indicates that \(\lambda_2^*(y) \leq \hat{\lambda}\) for any \(y \leq y_{\hat{\lambda}}\), and \(\lambda_2^*(y) > \hat{\lambda}\) for any \(y > y_{\hat{\lambda}}\), i.e., \(\lambda_2^*(\lambda)(y) = 0\) for any \(y \leq y_{\hat{\lambda}}\) and \(\lambda_2^*(\lambda)(y) > 0\) for any \(y > y_{\hat{\lambda}}\). Hence, \(y_{\hat{\lambda}}\) is a threshold point for the quantity-sales mode.

Next we show that \(\bar{y}_{\hat{\lambda}}\) is a threshold point for the unit-sales mode. We first show that there exists some \(\bar{\lambda} \in [0,m]\) such that \(\lambda - m\lambda_2^*(\lambda) > 0\) for any \(\lambda \in (0, \bar{\lambda}]\), and \(\lambda - m\lambda_2^*(\lambda) = 0\) for any \(\lambda \in [\bar{\lambda}, m]\). Recall the function \(\lambda_1^*(\lambda)\) defined in the proof of Proposition 5.1. The property \(\lambda_1^*(\lambda) + m\lambda_2^*(\lambda) = \lambda\) implies that it is
sufficient to show ρ* (λ) > 0 for any λ ∈ (0, X) and ρ* (λ) = 0 for any λ ∈ [X, m]. Let us define X as follows:

$$X = \begin{cases} \inf \{\lambda \in (0, m) : \frac{\partial R}{\partial \lambda_1}(0, \lambda) \leq 0\} & \text{if } \lambda \in (0, m), \ \frac{\partial R}{\partial \lambda_1}(0, \lambda) \neq 0, \\ m & \text{if } \lambda \in (0, m), \ \frac{\partial R}{\partial \lambda_1}(0, \lambda) = 0. \end{cases}$$

Assumption 3 yields that \( \frac{\partial R}{\partial \lambda_1}(0, \lambda) > 0 \) for any \( \lambda \in (0, X) \), and \( \frac{\partial R}{\partial \lambda_1}(0, \lambda) \leq 0 \) for any \( \lambda \in [X, m] \), which, combining with (5.5), shows that \( \rho^*_1(\lambda) > 0 \) for any \( \lambda \in (0, X) \), and \( \rho^*_1(\lambda) = 0 \) for any \( \lambda \in [X, m] \). Also note that \( \mathcal{C}(m) = \{0\} \), and hence \( \rho^*_1(m) = 0 \).

If \( y^*_t < \overline{y}_t \) and \( \rho^*_1(y^*_t) > 0 \), then the definition of \( \overline{y}_t \) indicates that \( \rho^*_1(y^*_t) - m \rho^*_2(\lambda^*_1(y^*_t)) > 0 \). The definition of \( \overline{y}_t \) implies that \( \rho^*_1(\overline{y}_t) - m \rho^*_2(\lambda^*_1(\overline{y}_t)) = 0 \). Therefore, \( \rho^*_1(\overline{y}_t) \geq X_t \). If \( y^*_t \geq \overline{y}_t \), Lemma 1 implies that \( \rho^*_1(y^*_t) \geq \rho^*_1(\overline{y}_t) \geq \overline{X}_t \). Hence, \( \rho^*_1(y^*_t) - m \rho^*_2(\lambda^*_1(y^*_t)) = 0 \), i.e., the firm should not quote the unit-sales price. □

**Infinite Horizon**

In this sub-section, we will show that the structural properties of the optimal policy for the finite horizon model also hold for the infinite horizon problem. Given the starting inventory level \( x_1 \), let \( V_{0,n}(x_1) \) denote the expected discounted profit from period 1 to period \( n \), i.e.,

$$V_{0,n}(x_1) = \max_{y_t(x_1) \geq x_1, \ 0 \leq \lambda_t(x_1) \leq m, y_t=1,...,n} E \left[ \sum_{t=1}^{n} \theta^{t-1} \left( -c(y_t(x_t) - x_t) - \ell(y_t(x_t) + \lambda_t(x_t)D_t - \varepsilon_t + \gamma(\lambda_t(x_t)) \right) \right].$$

(5.6)

Similar to the analysis in Section 2, we can also define \( W_{0,n}(x_1) = V_{0,n}(x_1) - cx_1 \). It can be proved that \( W_{0,n}(x) \) converges to a function \( W(x) \) such that

$$W(x) = \max_{y \geq x, \ 0 \leq \lambda \leq m} \left\{ (\theta - 1)c \gamma - E[\ell(y - \lambda D - \varepsilon)] + \gamma(\lambda) + \theta E[W(y - \lambda D - \varepsilon)] - \theta c \lambda \mu \right\},$$

(5.7)

and \( W(x) \) is the modified optimal profit function for the infinite horizon problem.

**Proposition 5.2.** There exists a concave function \( W(x) \) such that \( \lim_{n \to \infty} W_{0,n}(x) = W(x) \). Moreover, \( W(x) \) satisfies equation (5.7) and is the modified optimal profit function for the infinite horizon problem.

Proof. For any \( x, \lambda \in \mathbb{R} \) and \( n = 1, 2, ..., \), define \( U_n(x, \lambda) = \{(y, \lambda) \in [x, \infty) \times [0, m] : J_{0,n}(y, \lambda) \geq \Lambda \} \), where \( J_{0,n}(y, \lambda) = (\theta - 1)c \gamma - E[\ell(y - \lambda D - \varepsilon)] + \gamma(\lambda) + \theta E[W_{n-1}(y - \lambda D - \varepsilon)] - \theta c \lambda \mu \). Note that \( J_{0,n}(y, \lambda) \) is continuous with \( y \) and \( \lambda \), and \( \lim_{y \to \infty} J_{0,n}(y, \lambda) = -\infty \) for any fixed \( \lambda \) and \( n \). Therefore, \( U_n(x, \lambda) \) is a compact set in \([x, \infty) \times [0, \lambda_{max}]\). In accordance with Proposition 9.17 in Bertsekas and Shreve (1978), \( \lim_{n \to \infty} W_{0,n}(x) = W(x) \) and \( W(x) \) is the optimal profit function for the infinite horizon problem. The concavity of \( W(x) \) is due to the concavity of \( W_{0,n}(x) \), which can be obtained by the proof of Theorem 1. □

Let us define

$$J(y, \lambda) = (\theta - 1)c \gamma - E[\ell(y - \lambda D - \varepsilon)] + \gamma(\lambda) + \theta E[W(y - \lambda D - \varepsilon)] - \theta c \lambda \mu$$

and the maximizers \( \lambda^*(y) \) and \( S \) for \( J(y, \lambda) \) such that

$$\lambda^*(y) = \arg \max_{0 \leq \lambda \leq m} \left\{ J(y, \lambda) \right\} \quad \text{and} \quad S = \arg \max_y \left\{ J(y, \lambda^*(y)) \right\}.$$

(5.8)
Note that \( J(y, \lambda), \lambda^*(y) \) and \( S \) are the counterparts of \( J_t(y_t, \lambda_t), \lambda^*_t(y_t) \) and \( S_t \) defined in (14), (15) and (16), respectively. As \( J_\infty(y, \lambda) \) can be interpreted as the modified one-period profit, the myopic policy is somehow optimal. For a pure inventory problem, this has been shown in Chapter 3 of Heyman and Sobel (1984). We use the same spirit to prove the following theorem.

**Theorem 5.1.** If Assumption 1 holds, then the optimal inventory control policy can be described as follows. Let \( y^* \) denote the optimal order-up-to level for any period \( t = 1, 2, \ldots \). Then

\[
y^* = \begin{cases} 
  S & \text{if } x \leq S \\
  x & \text{if } x > S,
\end{cases}
\]

where \( S \) is defined in (5.9). Moreover, the value of \( S \) satisfies

\[
(S, \lambda^*(S)) = \arg\max_{y \in (-\infty, \infty), \lambda \in [0, m]} \left\{ J_\infty(y, \lambda) \right\},
\]

where the functions \( \lambda^*(y) \) and \( J_\infty(y, \lambda) \) are defined in (5.9) and (5.8), respectively.

**Proof.** Applying the concavity of \( W(x) \) shown in Proposition 5.2, the optimality of the base-stock policy can be established by the same argument used to prove Theorem 1. It remains to show that \( S \) and \( \lambda^*(S) \) correspond to the global maximizer of \( J_\infty(y, \lambda) \).

Consider any feasible policy \( y_t(x_t) \) and \( \lambda_t(x_t) \) taken in period \( t = 1, 2, \ldots \), i.e., the actions taken in each period \( t \), \( y_t(x_t) \) and \( \lambda_t(x_t) \) are functions of \( x_t \) such that \( y_t(x_t) \geq x_t \) and \( \lambda_t(x_t) \in [0, m] \). Without loss of generality, we set \( y_0(x_0) = x_1 \) and \( \lambda_0(x_0) = 0 \). Given the initial inventory level \( x_1 \), the expected discounted profit over the infinite horizon under the policy \( y_t(x_t) \) and \( \lambda_t(x_t) \) can be written as

\[
E \left[ \sum_{t=1}^{\infty} \theta^{t-1} \left( -c(y_t(x_t) - x_t) - \ell(y_t(x_t) - \lambda_t(x_t)D_t - \varepsilon_t) + \Gamma(\lambda_t(x_t)) \right) \right] = E \left[ \sum_{t=1}^{\infty} \theta^{t-1} \left( -cy_t(x_t) - \ell(y_t(x_t) - \lambda_t(x_t)D_t - \varepsilon_t) + \Gamma(\lambda_t(x_t)) \right) + \sum_{t=1}^{\infty} \theta^{t-1}cx_t \right]
\]

\[
= E \left[ \sum_{t=1}^{\infty} \theta^{t-1} \left( -cy_t(x_t) - \ell(y_t(x_t) - \lambda_t(x_t)D_t - \varepsilon_t) + \Gamma(\lambda_t(x_t)) \right) + \sum_{t=0}^{\infty} \theta^t c(x_t) \right]
\]

\[
= E \left[ \sum_{t=1}^{\infty} \theta^{t-1} \left( -cy_t(x_t) - \ell(y_t(x_t) - \lambda_t(x_t)D_t - \varepsilon_t) + \Gamma(\lambda_t(x_t)) \right) + \sum_{t=0}^{\infty} \theta^t c(y_t(x_t) - \lambda_t(x_t)D_t - \varepsilon_t) \right]
\]

\[
+ E[c(y_0(x_0) - \lambda_0(x_0)D_0)]
\]

\[
= cx_1 + \sum_{t=1}^{\infty} \theta^{t-1} E[J_\infty(y_t(x_t), \lambda_t(x_t))].
\]

The convergence of \( W_{0,n}(x) \) implies the convergence of \( V_{0,n}(x) \). Therefore, the optimal expected discounted profit in the infinite planning horizon can be written as

\[
V(x_1) = \lim_{n \to \infty} V_{0,n}(x_1) = cx_1 + \max_{y(x_t) \geq x_t, 0 \leq \lambda_t(x_t) \leq m} \left( \sum_{t=1}^{\infty} \theta^{t-1} E[J_\infty(y_t(x_t), \lambda_t(x_t))] \right).
\]
Obviously, \( W_{0,n}(x) = V_{0,n}(x) - cx \) implies that
\[
W(x_1) = V(x_1) - cx_1 = \max_{y(x_1) \geq x_1, 0 \leq \lambda(x_1) \leq m} \left( \sum_{t=1}^{\infty} \theta^{t-1} E[J_{\infty}(y_t(x_1), \lambda_t(x_1))] \right).
\] (5.10)

Let \( (S, \lambda^*(S)) = \arg\max_{y \in (-\infty, \infty), \lambda \in [0, m]} \{ J_{\infty}(y, \lambda) \} \). For each period, consider the strategy \( (S, \lambda^*(S)) \).

As \( \lambda D + \varepsilon \) is always nonnegative, the base stock level \( S \) must be greater than the initial inventory level \( S - \lambda^*(S) D - \varepsilon \). Therefore, (5.10) implies that \( (S, \lambda^*(S)) \) also maximizes \( E[W(y - \lambda D - \varepsilon)] \). Note that \( J(y, \lambda) = J_{\infty}(y, \lambda) + \theta E[W(y - \lambda D - \varepsilon)] \). We thus have \( (S, \lambda^*(S)) = \arg\max_{y \in (-\infty, \infty), \lambda \in [0, m]} \{ J(y, \lambda) \} \). \( \Box \)

Similar to the finite horizon model, the concavity of \( W(x) \) implies the supermodularity of \( J(y, \lambda) \), and thus \( \lambda^*(y) \) is also increasing in \( y \). Hence, the optimal pricing strategy for the infinite horizon model resembles its counterpart for the finite horizon model, which is Theorem 2. As the statement and proof are the same as those of Theorem 2, we omit them to avoid repetitions here.

**Proof of Proposition 1.** To simplify the notation, let us introduce \( \zeta(x) = xG^{-1}(1 - x) \). First, we establish the equivalence between the concavity of \( \zeta(x) \) and the increasing virtual value of \( V \). Note that
\[
\zeta'(x) = G^{-1}(1 - x) - xG^{-1}(1 - x) = G^{-1}(1 - x) - \frac{x}{g(G^{-1}(1 - x))},
\]
where the function \( g(\cdot) \) denotes the probability density function of \( V \). Let \( z = G^{-1}(1 - x) \), i.e., \( x = 1 - G(z) \). Then the first derivative of \( \zeta'(x) \) can be written as \( z = (1 - G(z))/g(z) = z - 1/h(z) \) where \( h(\cdot) \) denotes the failure rate of \( V \). According to Definition 1, \( \zeta'(x) \) is strictly increasing in \( z \) for any \( z \) in the support of \( V \), if and only if \( V \) has an increasing virtual value. Also, \( x \) is strictly decreasing in \( z \), and \( x \in [0, 1] \) if and only if \( z \) is in the support of \( V \). Therefore, the property that \( V \) has an increasing virtual value is equivalent to that \( \zeta'(x) \) is strictly decreasing in \( x \in [0, 1] \), i.e., \( \zeta(x) \) is strictly concave for any \( x \in [0, 1] \).

The expected one-period revenue can be written as
\[
R(\lambda_1, \lambda_2) = [\lambda_1 G^{-1}(1 - \lambda_1 - \lambda_2) + (a - 1)\lambda_2 G^{-1}(1 - \lambda_2) + (a - 1)\lambda_2 G^{-1}(1 - \lambda_1 - \lambda_2)] \cdot \mu
\]
\[
= \left[ ((\lambda_1 + \lambda_2)G^{-1}(1 - \lambda_1 - \lambda_2) + (a - 1)\lambda_2 G^{-1}(1 - \lambda_2)) \cdot \muight],
\]
where \( \mu = E[D] \).

The definition of \( R(\lambda_1, \lambda_2) \) in (5.12) and \( \zeta(x) = xG^{-1}(1 - x) \) yield that \( R(\lambda_1, \lambda_2) = [\zeta(\lambda_1 + \lambda_2) + (a - 1)\zeta(\lambda_2)] \cdot \mu \). Obviously, if \( \zeta(x) \) is strictly concave, then \( R(\lambda_1, \lambda_2) \) is also strictly concave. If \( \zeta(x) \) is not strictly concave, then \( R(\lambda_1, 0) = [\zeta(\lambda_1) + (a - 1)\zeta(0)] \cdot \mu \) is not strictly concave either. Therefore, Assumption 1 is satisfied if and only if \( \zeta(x) \) is strictly concave for any \( x \in [0, 1] \). Note that
\[
R_2(\lambda, \lambda) = R(\lambda - \lambda m, \lambda m) = (\zeta(\lambda m) + (a - 1)\zeta(0) \cdot \mu
\]
\[
R_1(\lambda_1, \lambda) = R\left(\lambda_1, \frac{\lambda_1 - \lambda}{m}\right) = \left[ (\lambda + (m - 1)\lambda_1) + (a - 1)\zeta\left(\frac{\lambda - \lambda_1}{m}\right) \right] \cdot \mu.
\]
(5.13)

Expression (5.13) implies that \( \frac{\partial}{\partial \lambda^2} R_2(\lambda, 0) = -(m - 1)\zeta'(\lambda) + (a - 1)\zeta'(0) \cdot \mu \) and \( \frac{\partial}{\partial \lambda} R_1(0, \lambda) = \frac{(m - a)\mu}{m} \zeta'(\frac{\lambda}{m}) \).

Therefore, the concavity of \( \zeta(x) \) in \([0, 1]\) is equivalent to that \( \frac{\partial}{\partial \lambda^2} R_2(\lambda, 0) \) is increasing in \( \lambda \in [0, 1] \) and \( \frac{\partial}{\partial \lambda} R_1(0, \lambda) \) is decreasing in \( \lambda \in [0, m] \), which yields Assumption 2 and Assumption 3. \( \Box \)
Proof of Proposition 2. Part (a). For convenience in notation, we define

\[ g_t(x) = (\theta c - c)x - E[\ell(x - \varepsilon_t)] + \theta E[W_{t+1}(x - \varepsilon_t)] \quad \text{and} \quad \bar{\Gamma}(x) = \Gamma(x) - \epsilon mx. \]

The dynamic programming equation can be written as

\[
W_t(x) = \max_{y \geq x, \lambda \in [0,m]} \left\{ (\theta c - c)E[y - \lambda D_t] - E[\ell(y - \lambda D_t - \varepsilon_t)] + \theta E[W_{t+1}(y - \lambda D_t - \varepsilon_t)] + \Gamma(\lambda) - \epsilon \mu \lambda \right\}
\]

\[
= \max_{y \geq x, \lambda \in [0,m]} \left\{ E[g_t(y - \lambda D_t)] + \bar{\Gamma}(\lambda) \right\}.
\]

Following (5.12), we have \( \Gamma(\lambda) = \left[ \zeta\left(\lambda - (m - 1)\lambda_2^*(\lambda)\right) + (a - 1)\zeta\left(\lambda_2^*(\lambda)\right) \right] \cdot \mu, \) where \( \zeta(x) = xG^{-1}(1 - x) \) and \( \lambda_2^*(\lambda) \) are defined in (12).

Let \( \bar{\lambda} = \arg\max_{x \in [0,1]} \{ \zeta(x) \} \). We want to show that \( \lambda_2^*(\bar{\lambda}) = \bar{\lambda}/m \) and \( \lambda - (m - 1)\lambda_2^*(\lambda) > 0 \) for any \( \lambda < \bar{\lambda}. \)

Since \( \zeta(x) \) is strictly concave, \( \Gamma(\lambda) \) is strictly concave and \( \bar{\lambda} \) is unique. Note that \( \Gamma(0) = 0 \) and \( \Gamma(\lambda) \) is strictly concave, then \( \lambda > 0. \) Define \( \tilde{\lambda} = \arg\max_{x \in [0,1]} \{ \zeta(x) \}. \) We have

\[
\Gamma(\tilde{\lambda}) = \left[ \zeta\left(\tilde{\lambda} - (m - 1)\lambda_2^*(\tilde{\lambda})\right) + (a - 1)\zeta\left(\lambda_2^*(\tilde{\lambda})\right) \right] \cdot \mu \leq a\zeta(\tilde{\lambda}) \cdot \mu. \tag{5.14}
\]

Note that \( a\zeta(\tilde{\lambda}) = \zeta(m\tilde{\lambda} - (m - 1)\lambda_2^*(\tilde{\lambda}) + (a - 1)\zeta(lambda_2^*(\tilde{\lambda})), \) hence

\[
a\zeta(\tilde{\lambda}) \cdot \mu \leq \max_{\lambda_2 \in \mathcal{S}(\lambda)} \left\{ \zeta(m\tilde{\lambda} - (m - 1)\lambda_2) + (a - 1)\zeta(lambda_2) \right\} \cdot \mu = \Gamma(m\tilde{\lambda}) \leq \Gamma(\tilde{\lambda}), \tag{5.15}
\]

where the last inequality is due to \( \tilde{\lambda} = \arg\max_{\lambda \in [0,m]} \{ \Gamma(\lambda) \}. \) Note that (5.14) with (5.15) imply that \( \Gamma(\tilde{\lambda}) = a\zeta(\tilde{\lambda}) \cdot \mu = \Gamma(m\tilde{\lambda}). \) As \( \zeta(x) \) and \( \Gamma(\lambda) \) are both strictly concave and \( a\zeta(\tilde{\lambda}) = \zeta(m\tilde{\lambda} - (m - 1)\tilde{\lambda} + (a - 1)\zeta(\tilde{\lambda}), \) we must have \( \tilde{\lambda} = m\tilde{\lambda} \) and \( \lambda_2(m\tilde{\lambda}) = \tilde{\lambda}, \) i.e., \( \lambda_2^*(\tilde{\lambda}) = \tilde{\lambda}/m. \) Next we show \( \lambda - m\lambda_2(\lambda) > 0 \) for any \( \lambda < m\tilde{\lambda}. \)

Since \( \zeta'(x) > 0 \) for \( x < \tilde{\lambda} \) and \( m > a, \) we can find an arbitrary small positive number \( \delta \) such that

\[
(m - 1) \cdot \frac{\zeta(x + (m - 1)\delta) - \zeta(x)}{(m - 1)\delta} > (a - 1) \cdot \frac{\zeta(x - \delta) - \zeta(x)}{-\delta}. \tag{5.16}
\]

After simplifying (5.16), we have \( \zeta(x + (m - 1)\delta) + (a - 1)\zeta(x - \delta) > a\zeta(x) \) for \( x < \tilde{\lambda}. \) Therefore, for any \( \lambda < m\tilde{\lambda} \) (i.e., \( \lambda/m < \tilde{\lambda} \), we can find an arbitrary small positive number \( \delta \) such that

\[
\zeta\left(\frac{\lambda}{m} + (m - 1)\delta\right) + (a - 1)\zeta\left(\frac{\lambda}{m} - \delta\right) > a\zeta\left(\frac{\lambda}{m}\right). \]

Because \( \frac{\lambda}{m} + (m - 1)\delta = \lambda - (m - 1)\left(\frac{\lambda}{m} - \delta\right), \) we have

\[
\zeta\left(\lambda - (m - 1)\left(\frac{\lambda}{m} - \delta\right)\right) + (a - 1)\zeta\left(\frac{\lambda}{m} - \delta\right) > a\zeta\left(\frac{\lambda}{m}\right). \]

Note that \( a\zeta\left(\frac{\lambda}{m}\right) = \zeta(\lambda - (m - 1)\lambda_2^*(\lambda) + (a - 1)\zeta(\lambda_2^*(\lambda)). \) Therefore, for given \( \lambda, \) the decision that \( \lambda_2 = \lambda/m \) is worse than the decision that \( \lambda_2 = \lambda/m - \delta, \) i.e., \( \lambda_2^*(\lambda) < \lambda/m \) for any \( \lambda < m\tilde{\lambda}. \) Thus, there must exist the unit-sales mode when \( 0 < \lambda < m\tilde{\lambda} = \tilde{\lambda}. \)

We define

\[
(S_t, \lambda_2^*) = \arg\max_{y, \lambda \in [0,m]} \left\{ E[g_t(y - \lambda D_t)] + \bar{\Gamma}(\lambda) \right\} \quad \text{and} \quad \bar{\lambda}_0 = \arg\max_{\lambda \in [0,m]} \{ \bar{\Gamma}(\lambda) \}. \tag{5.17}
\]
Given that \( \lambda_1^*(y) \) is an increasing function of \( y \), we prove \( S_t < \bar{y}_t \) by showing that \( \lambda_1^*(S_t) < \bar{\lambda} \).

Note that \( c > 0 \) and \( \bar{\Gamma}(\lambda) = \Gamma(x) - cxz \) is strictly concave, then \( \bar{\lambda}^0 = \arg\max_{\lambda \in [0, m]} \{ \bar{\Gamma}(\lambda) \} < \bar{\lambda} \). We first prove \( \lambda_1^* \leq \bar{\lambda}^0 \) by contradiction for the case of \( \bar{\lambda}^0 > 0 \). Suppose \( \lambda_1^* > \bar{\lambda}^0 \). Note that \( \bar{\lambda}^0 = \arg\max_{\lambda \in [0, m]} \{ \bar{\Gamma}(\lambda) \} \) and \( \bar{\Gamma}(\lambda) \) is strictly concave, we have

\[
\bar{\Gamma}(\lambda_1^*) < \bar{\Gamma}(\bar{\lambda}^0).
\]  

Define \( \bar{y}_t = \arg\max_y \{ g_t(y) \} \) and \( \bar{y}_t^0 \) that satisfies \( \frac{\bar{y}_t^0 - \bar{y}_t}{\bar{\lambda}^0} = \frac{S_t - \bar{y}_t}{\lambda_1^*} \), where \( \bar{y}_t \) denotes the minimum of all maximizers when the maximum of \( g_t(y) \) is not unique. As \( g_t(y) \) is concave and \( g_t(y) \) is increasing for \( y < \bar{y}_t \), (5.17) implies that \( S_t \geq \bar{y}_t \).

By \( \lambda_1^* > \bar{\lambda}^0 \) and \( \frac{\bar{y}_t^0 - \bar{y}_t}{\bar{\lambda}^0} = \frac{S_t - \bar{y}_t}{\lambda_1^*} \), we also have \( \bar{y}_t^0 < S_t \). Since \( S_t - \lambda_1^* z > \bar{y}_t^0 - \bar{\lambda}^0 z \) for \( z < \frac{\bar{y}_t^0 - \bar{y}_t}{\bar{\lambda}^0} \), \( S_t - \lambda_1^* z < \bar{y}_t^0 - \bar{\lambda}^0 z \) for \( z > \frac{\bar{y}_t^0 - \bar{y}_t}{\bar{\lambda}^0} \) and \( S_t - \lambda_1^* z = \bar{y}_t^0 - \bar{\lambda}^0 z \) for \( z = \frac{\bar{y}_t^0 - \bar{y}_t}{\bar{\lambda}^0} \) (i.e., \( z = \frac{S_t - \bar{y}_t}{\lambda_1^*} \)), we can derive \( g_t(S_t - \lambda_1^* z) \leq g_t(\bar{y}_t^0 - \bar{\lambda}^0 z) \) for \( z < \frac{\bar{y}_t^0 - \bar{y}_t}{\bar{\lambda}^0} \), \( g_t(S_t - \lambda_1^* z) \leq g_t(\bar{y}_t^0 - \bar{\lambda}^0 z) \) for \( z > \frac{\bar{y}_t^0 - \bar{y}_t}{\bar{\lambda}^0} \) and \( g_t(S_t - \lambda_1^* z) = g_t(\bar{y}_t^0 - \bar{\lambda}^0 z) \) for \( z = \frac{\bar{y}_t^0 - \bar{y}_t}{\bar{\lambda}^0} \) (i.e., \( z = \frac{S_t - \bar{y}_t}{\lambda_1^*} \)). This result is due to the fact that \( g_t(y) \) is increasing for \( y \leq \bar{y}_t \) and decreasing for \( y > \bar{y}_t \). Therefore, we have

\[
E[g_t(S_t - \lambda_1^* D_t)] = \int_0^{\frac{S_t - \bar{y}_t}{\lambda_1^*}} g_t(S_t - \lambda_1^* z) dF(z) + \int_{\frac{S_t - \bar{y}_t}{\lambda_1^*}}^{+\infty} g_t(S_t - \lambda_1^* z) dF(z)
\]
\[
\leq \int_0^{\frac{\bar{y}_t^0 - \bar{y}_t}{\lambda_1^*}} g_t(\bar{y}_t^0 - \bar{\lambda}^0 z) dF(z) + \int_{\frac{\bar{y}_t^0 - \bar{y}_t}{\lambda_1^*}}^{+\infty} g_t(\bar{y}_t^0 - \bar{\lambda}^0 z) dF(z) = E[g_t(\bar{y}_t^0 - \bar{\lambda}^0 D_t)].
\]  

Combining (5.18) with (5.19), we have

\[
E[g_t(S_t - \lambda_1^* D_t)] + \bar{\Gamma}(\lambda_1^*) < E[g_t(\bar{y}_t^0 - \bar{\lambda}^0 D_t)] + \bar{\Gamma}(\bar{\lambda}^0).
\]

Therefore, we obtain a contradiction with the definition of \( (S_t, \lambda_1^*) \) in (5.17). Hence \( \lambda_1^* \leq \bar{\lambda}^0 < \bar{\lambda} \).

If \( \bar{\lambda}^0 = 0 \), then from the definition of \( (S_t, \lambda_1^*) \) in (5.17), we have \( S_t = \bar{y}_t \) and \( \lambda_1^* = 0 < \bar{\lambda} \). This completes the proof of Part (a).

Part (b). Applying Theorems 2, the definition of \( \lambda_1^*(y) \) in (15), the definition of \( \lambda_2^*(\lambda) \) in (12) and the price functions in (6) lead to

\[
p_1^* = G^{-1} \left( 1 - \lambda_1^*(y_t^*) + (m - 1) \lambda_2^*(\lambda_1^*(y_t^*)) \right) \quad \text{for any } y_t^* < \bar{y}_t^*,
\]
\[
p_2^* = (a - 1)G^{-1} \left( 1 - \lambda_2^*(\lambda_1^*(y_t^*)) \right) + G^{-1} \left( 1 - \lambda_1^*(y_t^*) + (m - 1) \lambda_2^*(\lambda_1^*(y_t^*)) \right) \quad \text{for any } y_t^* > \bar{y}_t^*.
\]

Expression (5.13) implies that \( \frac{\partial^2}{\partial \lambda \partial \lambda} R_2(\lambda, \lambda_2) = -(m - 1)\mu c''(\lambda - (m - 1)\lambda_2) > 0 \), which yields that \( R_2(\lambda, \lambda_2) \) is supermodular and \( \lambda_2^*(\lambda) \) is increasing in \( \lambda \). Note that \( \lambda_1^*(y) \) is increasing function by Lemma 1. Also note that \( G^{-1}(x) \) is increasing as it is the inverse of the cumulative distribution function. Therefore, the function \( (a - 1)G^{-1}(1 - \lambda_2^*(\lambda_1^*(y_t^*))) \) is decreasing in \( y_t^* \), which makes it sufficient to show that \( \lambda_1^*(y_t^*) - (m - 1) \lambda_2^*(\lambda_1^*(y_t^*)) \) is increasing in \( y_t^* \). Recall that \( \lambda_1^*(y) \) is increasing in \( y \). The proposition would be proved if we could establish that \( \lambda - (m - 1) \lambda_2^*(\lambda) \) is increasing in \( \lambda \) for any \( \lambda \in [0, m] \).

Define \( \lambda_0 \) as \( \lambda_0 = \lambda_1 + \lambda_2 \). It is straightforward that \( \lambda_1 = \frac{m \lambda_0 - \lambda}{m - 1} \) and \( \lambda_2 = \frac{\lambda - \lambda_0}{m - 1} \), and hence the expected one period revenue can be represented as a function \( R_0(\lambda, \lambda_0) \) such that

\[
R_0(\lambda, \lambda_0) = R \left( \frac{m \lambda_0 - \lambda}{m - 1}, \frac{\lambda - \lambda_0}{m - 1} \right) = \left[ \zeta(\lambda_0) + (a - 1)\zeta \left( \frac{\lambda - \lambda_0}{m - 1} \right) \right] \cdot \mu
\]

27
where the equalities follow from (5.12) and \( \zeta(x) = xG^{-1}(1-x) \). Note that as shown in the proof of Proposition 1, \( \zeta(x) \) is a concave function in \([0, 1]\) if \( V \) has an increasing virtual value. According to the definition of \( \mathcal{A} \) in (2), we can show that the feasible region of \((\lambda, \lambda_0)\) is 

\[ \mathcal{D} = \left\{ (\lambda, \lambda_0) : \frac{m\lambda_0 - \lambda}{m-1} \geq 0, \ \frac{\lambda - \lambda_0}{m-1} \geq 0, \ \lambda_0 \leq 1 \right\} = \left\{ (\lambda, \lambda_0) : 0 \leq \lambda \leq m, \ \frac{\lambda}{m} \leq \lambda_0 \leq \min(\lambda, 1) \right\}. \]

Consider \( \lambda^*_0(\lambda) = \arg \max_{\lambda_0 \in \left[ \frac{\lambda}{m}, \min(\lambda, 1) \right]} \{ R_0(\lambda, \lambda_0) \} \). Using a similar argument in the proof of Proposition 5.1, we can prove that \( \lambda^*_0(\lambda) + (m-1)\lambda^*_2(\lambda) = \lambda \). Therefore, \( \lambda - (m-1)\lambda^*_2(\lambda) \) is increasing in \( \lambda \) as long as \( \lambda^*_0(\lambda) \) is increasing in \( \lambda \).

As the function \( \zeta(x) \) is concave for any \( x \in [0, 1] \), 
\[ \frac{\partial^2}{\partial \lambda \partial \lambda_0} R_0(\lambda, \lambda_0) = -\frac{s-1}{m-1} \mu \zeta'' \left( \frac{\lambda - \lambda_0}{m-1} \right) \geq 0, \]
which implies that \( R_0(\lambda, \lambda_0) \) is supermodular. Note that in the feasible set of \( \mathcal{D} \), both the lower bound and upper bound of \( \lambda_0 \) are increasing in \( \lambda \). Hence, we can show that \( \lambda^*_0(\lambda) \) is increasing in \( \lambda \). We thus complete the proof of monotonicity of \( p^*_1 \) and \( p^*_2 \).

Equation (6) and monotonicity of \( \lambda^*_2(\lambda^*_1(y_1^r)) \) imply that \( p^*_2 - p^*_1 \) always decreases with the optimal order up-to level \( y_1^r \).

**Proof of Proposition 3.** Recall that \( h(x) \) denotes the failure rate function of \( V \) and \([0, \overline{\gamma}] \) the support of \( V \). If \( V \) has an increasing failure rate, we must have 

\[ h(x) \geq \frac{a-1}{m-1} \times h \left( \frac{a-1}{m-1} x \right) \text{ for any } x \in [0, \overline{\gamma}] \]  

because \( a < m \). The proposition holds as long as 

\[ mp_1(\lambda - m\lambda^*_2(\lambda), \lambda^*_2(\lambda)) \geq p_2(\lambda - m\lambda^*_2(\lambda), \lambda^*_2(\lambda)) \]  

(5.21)  

for any \( \lambda \in [0, m] \), where \( \lambda^*_2(\lambda) \) is defined in (12), and the price functions \( p_1(\lambda_1, \lambda_2) \) and \( p_2(\lambda_1, \lambda_2) \) are shown in (6). Consider any \( \lambda \in [0, 1 - G \left( \frac{a-1}{m-1} \overline{\gamma} \right)] \). For any \( \lambda_2 \geq 0 \), (6) implies that 

\[ mp_1(\lambda - m\lambda_2, \lambda_2) - p_2(\lambda - m\lambda_2, \lambda_2) = (m-1)G^{-1}(1 - \lambda + (m-1)\lambda_2) - (a-1)G^{-1}(1 - \lambda_2). \]

Note that \( G^{-1}(1 - \lambda_2) \leq \overline{\gamma} \) as \( \lambda_2 \geq 0 \), and \( G^{-1}(1 - \lambda + (m-1)\lambda_2) \geq G^{-1} \left( G \left( \frac{a-1}{m-1} \overline{\gamma} \right) \right) = \frac{a-1}{m-1} \overline{\gamma} \), because \( G^{-1}(x) \) is increasing, \( \lambda \leq 1 - G \left( \frac{a-1}{m-1} \overline{\gamma} \right) \) and \( \lambda_2 \geq 0 \). It follows that \( mp_1(\lambda - m\lambda_2, \lambda_2) - p_2(\lambda - m\lambda_2, \lambda_2) \geq 0 \) for any \( \lambda_2 \geq 0 \), and hence (5.21) holds for any \( \lambda \in [0, 1 - G \left( \frac{a-1}{m-1} \overline{\gamma} \right)] \) as \( \lambda^*_2(\lambda) \geq 0 \).

Next, we prove inequality (5.21) for any \( \lambda \in \left[ 1 - G \left( \frac{a-1}{m-1} \overline{\gamma} \right), m \right] \). Let us define the function \( \phi(\lambda_2) \) such that 

\[ \phi(\lambda_2) = 1 + (m-1)\lambda_2 - G \left( \frac{a-1}{m-1} G^{-1}(1 - \lambda_2) \right). \]  

(5.22)  

Since \( G(x) \) is a cumulative distribution function, it is straightforward to show that \( \phi(\lambda_2) \) is strictly increasing in \( \lambda_2 \in [0, 1] \). Therefore, the inverse function of \( \phi(\lambda_2) \), \( \phi^{-1}(\lambda) \), is well-defined for any \( \lambda \in [\phi(0), \phi(1)] \), i.e., 

\[ \lambda \in \left[ 1 - G \left( \frac{a-1}{m-1} \overline{\gamma} \right), m \right]. \]

For any \( \lambda_2 \in [0, 1] \), we obtain \( \phi(\lambda_2) \leq 1 + (m-1)\lambda_2 \) because \( G \left( \frac{a-1}{m-1} G^{-1}(1 - \lambda_2) \right) \geq 0 \), and \( \phi(\lambda_2) \geq m\lambda_2 \) because \( G \left( \frac{a-1}{m-1} G^{-1}(1 - \lambda_2) \right) \geq G \left( G^{-1}(1 - \lambda_2) \right) = 1 - \lambda_2 \), where the inequality holds because of the monotonicity of \( G(x) \). The definitions of \( \mathcal{D} \) in (10) and \( \mathcal{A} \) in (2) imply that \( \mathcal{D} = \{(\lambda, \lambda_2) : 0 \leq \lambda \leq 1, \ m\lambda_2 \leq \]
\( \lambda \leq 1 + (m - 1)\lambda_2 \). Therefore, we have \((\phi(\lambda_2), \lambda_2) \in \mathcal{B}\) for any \(\lambda_2 \in [0, 1]\), i.e., \((\lambda_2, \phi^{-1}(\lambda_2)) \in \mathcal{B}\) for any \(\lambda \in [\phi(0), \phi(1)] = [1 - G\left(\frac{\alpha - 1}{m-1}\right), m]\). Applying the definition of \(\mathcal{B}(\lambda)\) in (11), we have \(\phi^{-1}(\lambda) \in \mathcal{B}(\lambda)\) for any \(\lambda \in [1 - G\left(\frac{\alpha - 1}{m-1}\right), m]\). Note that \(\mathcal{B}(\lambda)\) is the feasible region of the maximization problem in (12), whose optimal solution is \(\lambda_2^*(\lambda)\). Consider the first partial derivative of \(R_2(\lambda, \lambda_2)\) with respect to \(\lambda_2\) at any point \((\phi(\lambda_2), \lambda_2)\) where \(\lambda_2 \in [0, 1]\). According to (5.13) and (5.11), we have

\[
\frac{\partial}{\partial \lambda_2} R_2(\phi(\lambda_2), \lambda_2) = \left( a - 1 \right) \left( G^{-1}(1 - \lambda_2) - \frac{\lambda_2}{g(G^{-1}(1 - \lambda_2))} \right) - (m - 1) \left( G^{-1}(1 - \phi(\lambda_2) + (m - 1)\lambda_2) - \frac{\phi(\lambda_2) - (m - 1)\lambda_2}{g(G^{-1}(1 - \phi(\lambda_2) + (m - 1)\lambda_2))} \right),
\]

where \(g(x)\) denotes the probability density function of \(V\). The definition of \(\phi(\lambda_2)\) yields that

\[
\frac{\partial}{\partial \lambda_2} R_2(\phi(\lambda_2), \lambda_2) = \left( a - 1 \right) \left( G^{-1}(1 - \lambda_2) - \frac{\lambda_2}{g(G^{-1}(1 - \lambda_2))} \right) - (m - 1) \left( 1 - G\left(\frac{\alpha - 1}{m-1}\right) \right) \frac{a - 1}{g\left(\frac{\alpha - 1}{m-1}\right)} \left( G\left(\frac{\alpha - 1}{m-1}\right) \right) - \frac{(a - 1)\lambda_2}{g(G^{-1}(1 - \lambda_2))} \right),
\]

where the inequality follows from the condition (5.20) and \(G^{-1}(1 - \lambda_2) \in [0, \pi]\) because \(\lambda_2 \in [0, 1]\). Applying the definition of \(\phi^{-1}(\lambda)\), we obtain for any \(\lambda \in [1 - G\left(\frac{\alpha - 1}{m-1}\right), m]\),

\[
\frac{\partial}{\partial \lambda_2} R_2(\lambda, \phi^{-1}(\lambda)) \geq 0.
\] (5.23)

Recall that \(\phi^{-1}(\lambda) \in \mathcal{B}(\lambda)\) and the function \(R_2(\lambda, \lambda_2)\) is concave. The definition of \(\lambda_2^*(\lambda)\) in (12) implies that \(\lambda_2^*(\lambda) \geq \phi^{-1}(\lambda)\) for any \(\lambda \in [1 - G\left(\frac{\alpha - 1}{m-1}\right), m]\). For any \(\lambda_2 \in [0, 1]\) and \(\lambda \leq \phi(\lambda_2)\), we have

\[
mp_1(\lambda - m\lambda_2, \lambda_2) - p_2(\lambda - m\lambda_2, \lambda_2)
\]

\[
= (m - 1)G^{-1}(1 - \lambda + (m - 1)\lambda_2) - (a - 1)G^{-1}(1 - \lambda_2)
\]

\[
\geq (m - 1)G^{-1}(1 - \phi(\lambda_2) + (m - 1)\lambda_2) - (a - 1)G^{-1}(1 - \lambda_2)
\]

\[
= (m - 1)G^{-1}\left( G\left(\frac{a - 1}{m - 1}\right) \right) G^{-1}(1 - \lambda_2) - (a - 1)G^{-1}(1 - \lambda_2) = 0,
\]

where the inequality follows from the monotonicity of \(G^{-1}(x)\). Hence, \(mp_1(\lambda - m\lambda_2, \lambda_2) - p_2(\lambda - m\lambda_2, \lambda_2) \geq 0\), for any \(\lambda \in [1 - G\left(\frac{\alpha - 1}{m-1}\right), m]\) and \(\lambda_2 \geq \phi^{-1}(\lambda)\). Recall that we have shown \(\lambda_2^*(\lambda) \geq \phi^{-1}(\lambda)\) for any \(\lambda \in [1 - G\left(\frac{\alpha - 1}{m-1}\right), m]\). Therefore, (5.21) holds for any \(\lambda \in [1 - G\left(\frac{\alpha - 1}{m-1}\right), m]\). \(\Box\)

**Proof of Proposition 4.** Note that both unit-sales and quantity-sales prices exist only when \(0 < \lambda_2^*(\lambda) < \frac{A}{m}\) (otherwise, either \(\lambda_2^*(\lambda) = 0\) or \(\lambda_2^*(\lambda) = \frac{A}{m} = \lambda_2^*(\lambda) = 0\)). First, we wish to establish that if the failure rate function \(h(x)\) of \(V\) satisfies (5.20), then

\[
\frac{\partial}{\partial \lambda_2} R_2(\lambda, \lambda_2^*(\lambda)) = 0 \text{ for any } \lambda \text{ such that } 0 < \lambda_2^*(\lambda) < \frac{A}{m},
\] (5.24)

29
where $\lambda_2^*(\lambda)$ is defined in (12). This property is proved by considering the following three cases.

Case 1. Suppose that $0 \leq \lambda \leq 1$. According to the definition of $\mathcal{A}$ in (2), the set $\mathcal{B}$ in (10) is

$$\mathcal{B} = \left\{ (\lambda, \lambda_2) : 0 \leq \lambda \leq m, \max \left( 0, \frac{\lambda - 1}{m - 1} \right) \leq \lambda_2 \leq \frac{\lambda}{m} \right\}.$$  

Note that (11) implies $\mathcal{B}(\lambda)$ to be the interval $[0, \frac{\lambda}{m}]$ for any $\lambda \in [0, 1]$. Recall that $\lambda_2^*(\lambda)$ is the maximizer of $R_2(\lambda, \lambda_2)$ for any $\lambda_2 \in \mathcal{B}(\lambda)$. The concavity of $R_2(\lambda, \lambda_2)$ implies $\frac{\partial}{\partial \lambda_2} R_2(\lambda, \lambda_2^*(\lambda)) = 0$ for any $\lambda \in [0, 1]$ such that $0 < \lambda_2^*(\lambda) < \frac{\lambda}{m}$.

Case 2. Suppose that $\lambda \in (1, m)$ and $0 < \lambda_2^*(\lambda) < \frac{\lambda}{m}$. $\lambda > 1$ implies that $\lambda > 1 - G(\frac{a - 1}{m - 1} \tau)$, where $\tau$ denotes the upper limit of $V$’s support. Under the condition (5.20), we obtained in the proof of Proposition 3 that $\lambda_2^*(\lambda) \geq \phi^{-1}(\lambda)$ and $\frac{\partial}{\partial \lambda_2} R_2(\lambda, \phi^{-1}(\lambda)) \geq 0$, where $\phi(\lambda)$ is defined in (5.22). If $\lambda_2^*(\lambda) > \phi^{-1}(\lambda)$, then the definition of $\phi(x)$ in (5.22) implies that

$$\lambda < \phi(\lambda_2^*(\lambda)) = 1 + (m - 1)\lambda_2^*(\lambda) - G \left( \frac{a - 1}{m - 1} G^{-1}(1 - \lambda_2^*(\lambda)) \right) \leq 1 + (m - 1)\lambda_2^*(\lambda),$$

where the second inequality is obtained from the non-negativity of $G(x)$, and hence we have $\lambda_2^*(\lambda) > \frac{\lambda - 1}{m - 1}$. Since $0 < \lambda_2^*(\lambda) < \frac{\lambda}{m}$, it yields that $\frac{\lambda - 1}{m - 1} < \lambda_2^*(\lambda) < \frac{\lambda}{m}$. Also note that $\mathcal{B}(\lambda)$ is the interval $\left[ \frac{\lambda - 1}{m - 1}, \frac{\lambda}{m} \right]$. Therefore, the concavity of $R_2(\lambda, \lambda_2)$ and the definition of $\lambda_2^*(\lambda)$ in (12) imply that $\frac{\partial}{\partial \lambda_2} R_2(\lambda, \lambda_2^*(\lambda)) = 0$. If $\lambda_2^*(\lambda) = \phi^{-1}(\lambda)$, then (5.23) leads to

$$\frac{\partial}{\partial \lambda_2} R_2(\lambda, \lambda_2^*(\lambda)) = \frac{\partial}{\partial \lambda_2} R_2(\lambda, \phi^{-1}(\lambda)) \geq 0. \quad (5.25)$$

Recall that $\lambda_2^*(\lambda) < \frac{\lambda}{m}$ and $\mathcal{B}(\lambda)$ is the interval $\left[ \frac{\lambda - 1}{m - 1}, \frac{\lambda}{m} \right]$. We obtain from the concavity of $R_2(\lambda, \lambda_2)$ and the definition of $\lambda_2^*(\lambda)$ that $\frac{\partial}{\partial \lambda_2} R_2(\lambda, \lambda_2^*(\lambda)) \leq 0$, which, combining with (5.25), yields $\frac{\partial}{\partial \lambda_2} R_2(\lambda, \lambda_2^*(\lambda)) = 0$.

Case 3. If $\lambda = m$, then $(\lambda - 1)/(m - 1) = \frac{\lambda}{m} = 1$, i.e., $\mathcal{B}(\lambda) = \{1\}$. The definition of $\lambda_2^*(\lambda)$ in (12) leads to $\lambda_2^*(\lambda) = 1 = \frac{\lambda}{m}$, which violates the condition that $0 < \lambda_2^*(\lambda) < \frac{\lambda}{m}$.

Similar to Proposition 1, let $\zeta(x) = xG^{-1}(1 - x)$. According to (5.13), we have

$$\frac{\partial}{\partial \lambda_2} R_2(\lambda, \lambda_2) = \left[ -(m - 1)\zeta'(\lambda - (m - 1)\lambda_2) + (a - 1)\zeta'(\lambda_2) \right] \cdot \mu.$$  

Recall that we proved in Proposition 1 that $\zeta'(x)$ is strictly decreasing in $x$ if $V$ has an increasing virtual value, and hence the inverse function of $\zeta'(x)$ exists, which is denoted by $\zeta^{-1}(x)$. Applying this notation, (5.24) yields that

$$\lambda - (m - 1)\lambda_2^*(\lambda) = \zeta'^{-1} \left( \frac{a - 1}{m - 1} \zeta'(\lambda_2^*(\lambda)) \right) \text{ for any } \lambda \text{ such that } 0 < \lambda_2^*(\lambda) < \frac{\lambda}{m}. \quad (5.26)$$

Moreover, $0 < \lambda_2^*(\lambda) < \frac{\lambda}{m}$ also implies that $\lambda - (m - 1)\lambda_2^*(\lambda) > \lambda_2^*(\lambda)$ and hence $\zeta'^{-1} \left( \frac{a - 1}{m - 1} \zeta'(\lambda_2^*(\lambda)) \right) > \lambda_2^*(\lambda)$. Note that $\zeta(x)$ is strictly concave, and hence $\zeta'(x)$ is strictly decreasing. Therefore, we obtain $\frac{a - 1}{m - 1} \zeta'(\lambda_2^*(\lambda)) < \zeta'(\lambda_2^*(\lambda))$. Since $a < m$, it yields that

$$\zeta'(\lambda_2^*(\lambda)) > 0 \text{ for any } \lambda \text{ such that } 0 < \lambda_2^*(\lambda) < \frac{\lambda}{m}. \quad (5.27)$$
Consider any \( y^*_t \in (y_t, \bar{y}_t) \). The definitions of \( \lambda^*_t(y) \) in (15), the definition of \( \lambda^*_2(\lambda) \) in (12) and the price functions in (6) means

\[
mp^*_t - p^*_t = mp_1 \left( \lambda^*_t(y^*_t) - m\lambda^*_2(\lambda^*_t(y^*_t)) \right) - p_2 \left( \lambda^*_t(y^*_t) - m\lambda^*_2(\lambda^*_t(y^*_t)) \right) = (m - 1)G^{-1}(1 - \lambda^*_t(y^*_t) + (m - 1)\lambda^*_2(\lambda^*_t(y^*_t))) - (a - 1)G^{-1}(1 - \lambda^*_2(\lambda^*_t(y^*_t))).
\]

Since \( y_t < y^*_t < \bar{y}_t \), Theorems 2 implies that \( 0 < \lambda^*_2(\lambda^*_t(y^*_t)) < \frac{\lambda^*_2(y^*_t)}{m} \). Inserting (5.26) into above equality, we obtain

\[
mp^*_t - p^*_t = (m - 1)G^{-1} \left( 1 - \zeta'^{-1} \left( \frac{a - 1}{m - 1} \zeta'(\lambda^*_2(\lambda^*_t(y^*_t))) \right) \right) - (a - 1)G^{-1}(1 - \lambda^*_2(\lambda^*_t(y^*_t))). \tag{5.28}
\]

Recall that \( \lambda^*_t(y) \) is increasing in \( y \) and \( \lambda^*_2(\lambda) \) is increasing in \( \lambda \) (see Lemma 1 and the proof of Proposition 2, respectively). Therefore, the monotonicity of \( mp^*_t - p^*_t \) is equivalent to the monotonicity of the following function \( \delta(\lambda_2) \):

\[
\delta(\lambda_2) = (m - 1)G^{-1} \left( 1 - \zeta'^{-1} \left( \frac{a - 1}{m - 1} \zeta'(\lambda^*_2(\lambda^*_t(y^*_t))) \right) \right) - (a - 1)G^{-1}(1 - \lambda^*_2(\lambda^*_t(y^*_t))),
\]

whose domain is \( \{\lambda_2 \in [0, 1] : \zeta'(\lambda_2) > 0\} \) by (5.27). In other words, Proposition 4 would be proved if we could show that (i) \( \delta(\lambda_2) \) is increasing if \( 1/h(x) \) is convex, and (ii) \( \delta(\lambda_2) \) is decreasing if \( 1/h(x) \) is concave.

Consider the first derivative of the function \( \delta(\lambda_2) \), i.e.,

\[
\delta'(\lambda_2) = -(m - 1)G^{-1}(1 - \lambda^*_2) \times \zeta'^{-1} \left( \frac{a - 1}{m - 1} \zeta' \right) + (a - 1)G^{-1}(1 - \lambda_2) \times \zeta'^{-1} \left( \frac{a - 1}{m - 1} \zeta' \right) \times \zeta''(\lambda_2) \times \zeta'^{-1} \left( \frac{a - 1}{m - 1} \zeta' \right) + (a - 1)G^{-1}(1 - \lambda_2).
\]

Obviously, \( \delta(\lambda_2) \) is monotonically increasing iff \( \delta'(\lambda_2) \geq 0 \). According to the strict concavity of \( \zeta(x) \), i.e., \( \zeta''(x) < 0 \), \( \delta'(\lambda_2) \geq 0 \) is equivalent to \( \frac{G^{-1}(1 - \lambda_2)}{\zeta'(\lambda_2)} - \frac{G^{-1}(1 - \lambda^*_2)}{\zeta'(\lambda^*_2)} \leq 0 \). Recall that the domain of \( \delta(\lambda_2) \) is \( \{\lambda_2 \in [0, 1] : \zeta'(\lambda_2) > 0\} \). Therefore, applying the strict concavity of \( \zeta(x) \) and the fact that \( a < m \), we obtain \( \zeta'^{-1} \left( \frac{a - 1}{m - 1} \zeta'(\lambda_2) \right) > \lambda_2 \). As a result, a sufficient condition for \( \delta(\lambda_2) \) and thus \( mp^*_t - p^*_t \) to be monotone increasing is that \( \frac{G^{-1}(1 - x)}{\zeta''(x)} \) is increasing for any \( x \in [0, 1] \).

By the same argument, \( \delta(\lambda_2) \), and hence \( mp^*_t - p^*_t \), decreases if \( \frac{G^{-1}(1 - x)}{\zeta''(x)} \) is decreasing for any \( x \in [0, 1] \).

Note that \( G^{-1}(1 - x) = \frac{1}{g(G^{-1}(1 - x))} \) and \( \zeta''(x) = xG^{-1}(1 - 2G^{-1}(1 - x)) = -\frac{sx'G^{-1}(1 - x)}{g(G^{-1}(1 - x))} - \frac{2G^{-1}(1 - x)}{g(G^{-1}(1 - x))} \). Therefore, \( \frac{G^{-1}(1 - x)}{\zeta''(x)} = \frac{xG^{-1}(1 - x)}{g(G^{-1}(1 - x))} - \frac{2G^{-1}(1 - x)}{g(G^{-1}(1 - x))} \). Let \( z = G^{-1}(1 - x) \), i.e., \( x = 1 - G(z) \). If follows directly that

\[
\frac{G^{-1}(1 - x)}{\zeta''(x)} = \frac{1}{1 - G(z)G'(z)} = \frac{z}{(g(z))^2} - 2. \tag{5.29}
\]

Consider the function \( 1/h(x) \), where \( h(x) = g(x)/(1 - G(x)) \) is the failure rate of \( V \). Obviously,

\[
\frac{d}{dx} \left( \frac{1}{h(x)} \right) = \left( \frac{1 - G(x)g'(x)}{(g(x))^2} - \frac{g(x)g'(x)}{(g(x))^2} \right) = \frac{(1 - G(x))g'(x)}{(g(x))^2} - 1.
\]
If $1/h(x)$ is convex (concave), $-(1-G(x))g'(x)$ is increasing (decreasing). According to (5.29), $\frac{G^{-1}(1-x)}{g'(x)}$ is decreasing (increasing) in $z$, and hence it is increasing (decreasing) in $x$ as $x = 1 - G(z)$ is decreasing in $z$, which yields that both $\delta(\lambda_2)$ and $mp^*_t - p^*_2t$ are increasing (decreasing).

**Proof of Lemma 2.** Before proceeding to the proof, we develop a complete formulation of the infinite-horizon joint inventory and pricing model in which only a unit price is quoted. Given the initial inventory $x$, the optimal expected profit for this problem can be expressed as

$$\hat{V}(x) = cx + \hat{W}(x)$$

where $\hat{W}(x)$ satisfies

$$\hat{W}(x) = \max_{y \geq x} \left\{ (\theta - 1)cy - \mathbb{E}[\ell(y - \lambda D - \varepsilon)] + \hat{\Gamma}(\lambda) + \theta \mathbb{E}[\hat{W}(y - \lambda D - \varepsilon)] - \theta c \lambda \mu \right\}.$$

It is easy to see that the structural properties of the optimal policy are also applicable to this model, e.g., a base-stock inventory control policy is optimal and the optimal value of $\lambda$ is increasing the optimal order-up-to level. Clearly, we have $V(x) \geq \hat{V}(x)$ for any $x$.

Consider the function $J_\infty(y, \lambda)$ defined in (5.8), i.e.,

$$J_\infty(y, \lambda) = (\theta - 1)cy - \mathbb{E}[\ell(y - \lambda D - \varepsilon)] + \Gamma(\lambda) - \theta c \lambda \mu$$

where

$$\Gamma(\lambda) = \max_{\lambda_0} \max_{\frac{\lambda - \lambda_0}{\lambda_0 - \lambda} \leq \lambda_2 \leq \frac{\lambda}{\lambda_0}} \left\{ [(\lambda - (m - 1)\lambda_2)G^{-1}(1 - \lambda + (m - 1)\lambda_2) + (a - 1)\lambda_2G^{-1}(1 - \lambda_2)] \cdot \mu \right\}. \quad (5.31)$$

According to Theorem 5.1, the base-stock level to the infinite-horizon model with quantity discount is $S$ and the corresponding optimal value of $\lambda$ is $\lambda^*(S) = \lambda^\circ$. Therefore, for any $x \leq S$,

$$V(x) = cx + \sum_{t=1}^{\infty} \theta^{t-1} J_\infty(S, \lambda^\circ) = cx + \frac{J_\infty(S, \lambda^\circ)}{1 - \theta}.$$ 

It follows that $V(0) = \frac{J_\infty(S, \lambda^\circ)}{1 - \theta}$. Following the same argument, we complete the proof by showing that

$$\hat{V}(0) = \frac{J_\infty(S, \lambda^\circ)}{1 - \theta} \text{ and } \Delta = \frac{J_\infty(S, \lambda^\circ) - J_\infty(S, \lambda^\circ)}{1 - \theta}. \quad \square$$

**Proof of Proposition 5.** We first need to show that $\Gamma'(\lambda) \geq \hat{\Gamma}'(\lambda)$ for any $\lambda \in [0, 1]$. Note that $\lambda_2^*(\lambda)$ defined in (12) corresponds to the optimal solution to the optimization problem in (5.31). Therefore, we have

$$\Gamma(\lambda) = \hat{\Gamma}(\lambda - (m - 1)\lambda_2^*(\lambda)) + (a - 1)\hat{\Gamma}(\lambda_2^*(\lambda)). \quad (5.32)$$

Consider the following three cases.

Case 1. Suppose that $\lambda \leq \lambda$, where $\lambda$ is defined in (5.1). Note that the definition of $\lambda$ yields $\lambda \leq 1$, because (5.31) implies that $\lambda_2^*(\lambda) > 0$ for any $\lambda > 1$. For any $\lambda \leq \lambda$, the definition of $\lambda$ also implies that $\lambda_2^*(\lambda) = 0$, and hence (5.32) yields that $\Gamma(\lambda) = \lambda G^{-1}(1 - \lambda) \cdot \mu = \hat{\Gamma}(\lambda)$. It follows that $\Gamma'(\lambda) \geq \hat{\Gamma}'(\lambda)$.

Case 2. Suppose $\lambda < \lambda \leq 1$ and $\lambda < \lambda$, where $\lambda$ was defined in the proof of Theorem 2. According to the monotonicity of $\lambda_2^*(\lambda)$ shown in the proof Proposition 2, the definition of $\lambda$ implies that $\lambda_2^*(\lambda) > 0$ since $\lambda < \lambda$. 

32
We also have $\lambda_2^*(\lambda) < \frac{\lambda}{m}$ as $\lambda < \overline{\lambda}$. Recall that $\lambda \in [0,1]$, and hence

$$\Gamma(\lambda) = \max_{0 \leq \lambda_2 \leq \lambda} \{R_2(\lambda, \lambda_2)\} = \max_{0 \leq \lambda_2 \leq \lambda} \left\{\hat{\Gamma}((\lambda - (m-1)\lambda_2) + (a-1)\hat{\Gamma}(\lambda_2))\right\}, \quad (5.33)$$

where the objective function is concave and the feasible region is $[0, \frac{\lambda}{m}]$. Therefore, applying the first order condition, the property that $\lambda_2^*(\lambda) \in (0, \frac{\lambda}{m})$ yields that \( \frac{\partial}{\partial \lambda_2} R_2(\lambda, \lambda_2(\lambda)) = 0 \), i.e.,

$$-(m-1)\hat{\Gamma}'((\lambda - (m-1)\lambda_2^*(\lambda)) + (a-1)\hat{\Gamma}'(\lambda_2^*(\lambda)) = 0. \tag{5.34}$$

Taking derivative of (5.32) gives

$$\Gamma'(\lambda) = (1 - (m-1)\lambda_2^*(\lambda))\hat{\Gamma}'((\lambda - (m-1)\lambda_2^*(\lambda)) + (a-1)\lambda_2^*(\lambda))\hat{\Gamma}'(\lambda_2^*(\lambda))$$

$$= \hat{\Gamma}'(\lambda - m(1)\lambda_2^*(\lambda)) + \lambda_2^*(\lambda)((m-1)\hat{\Gamma}'((\lambda - (m-1)\lambda_2^*(\lambda)) + (a-1)\hat{\Gamma}'(\lambda_2^*(\lambda))$$

$$= \hat{\Gamma}'(\lambda - (m-1)\lambda_2^*(\lambda)),$$

where the last equality follows from (5.34). Note that $\lambda_2^*(\lambda) \in (0, \frac{\lambda}{m})$, and hence $0 < \lambda - (m-1)\lambda_2^*(\lambda) < \lambda$. According to the concavity of $\hat{\Gamma}(\lambda)$, we have $\hat{\Gamma}'(\lambda - (m-1)\lambda_2^*(\lambda)) \geq \hat{\Gamma}'(\lambda)$, which yields that $\Gamma'(\lambda) \geq \hat{\Gamma}'(\lambda)$.

Case 3. Suppose that $\underline{\lambda} < \lambda \leq 1$ and $\lambda \geq \overline{\lambda}$. By definition of $\overline{\lambda}$, we have $\lambda_2^*(\lambda) = \frac{\lambda}{m}$, and hence (5.32) shows that $\Gamma(\lambda) = a\hat{\Gamma}((\frac{\lambda}{m})$ and $\Gamma'(\lambda) = \frac{a}{m} \hat{\Gamma}'\left((\frac{\lambda}{m})\right)$. Again, as $\lambda \in [0,1]$, the function $\Gamma(\lambda)$ can be written in the form of (5.33). Since $\lambda_2^*(\lambda) = \frac{\lambda}{m}$, the concavity of $R_2(\lambda, \lambda_2)$ yields that $\frac{\partial}{\partial \lambda_2} R_2(\lambda, \frac{\lambda}{m}) \geq 0$, i.e.,

$$-(m-1)\hat{\Gamma}'(\frac{\lambda}{m}) + (a-1)\hat{\Gamma}'(\frac{\lambda}{m}) = (a-m)\hat{\Gamma}'(\frac{\lambda}{m}) \geq 0,$$

which implies $\hat{\Gamma}'(\frac{\lambda}{m}) \leq 0$. Recall that $m > a$, we obtain $\Gamma'(\lambda) = a\hat{\Gamma}'(\frac{\lambda}{m}) \geq \hat{\Gamma}'(\frac{\lambda}{m}) \geq \hat{\Gamma}'(\lambda)'$, where the inequality follows from the concavity of $\hat{\Gamma}(\lambda)$ and the property that $\frac{\lambda}{m} \leq \lambda$.

This completes the proof $\Gamma'(\lambda) \geq \hat{\Gamma}'(\lambda)$. Similar to the definition of $\lambda^*(y)$ in (5.9), we let

$$\hat{\lambda}^*(y) = \arg\max_{\lambda \in [0,1]} \{J_{\infty}(y, \lambda)\}. \tag{5.35}$$

Then $\lambda^*(S) = \lambda^o$ and $\hat{\lambda}^*(\hat{S}) = \hat{\lambda}^o$. We show that $\lambda^*(y) \geq \hat{\lambda}^*(y)$ by considering the following three subcases.

Subcase 1. If $\hat{\lambda}^*(y) = 0$, then it is trivial that $\lambda^*(y) \geq \hat{\lambda}^*(y)$ since $\lambda^*(y) \geq 0$.

Subcase 2. Suppose that $\hat{\lambda}^*(y) \in (0, 1)$. The first order condition shows that $\text{E}[\ell'(y - \hat{\lambda}^*(y)D - \epsilon)D] + \hat{\Gamma}'(\hat{\lambda}^*(y)) - \theta c \mu = 0$. As $\text{E}[\ell'(y - \lambda D - \epsilon)D] + \Gamma'(\lambda)$ is decreasing in $\lambda$, we have $\{\text{E}[\ell'(y - \lambda D - \epsilon)D] + \hat{\Gamma}'(\lambda) - \theta c \mu = 0\}_{\lambda=0} > 0$. Note that $\Gamma'(\lambda) \geq \hat{\Gamma}'(\lambda)$, then $\{\text{E}[\ell'(y - \lambda D - \epsilon)D] + \Gamma'(\lambda) - \theta c \mu = 0\}_{\lambda=0} > 0$ and $\lambda^*(y) > 0$. If $\lambda^*(y) = m$, it is trivial that $\lambda^*(y) \geq \hat{\lambda}^*(y)$. If $\lambda^*(y) < m$, the first order condition shows that $\text{E}[\ell'(y - \lambda^*(y)D - \epsilon)D] + \Gamma'(\lambda^*(y)) - \theta c \mu = 0$. As $\Gamma'(\lambda) \geq \hat{\Gamma}'(\lambda)$ and $\text{E}[\ell'(y - \lambda D - \epsilon)D] + \Gamma'(\lambda)$ is decreasing in $\lambda$, the first order conditions satisfied by $\hat{\lambda}^*(y)$ and $\lambda^*(y)$ derive $\lambda^*(y) \geq \hat{\lambda}^*(y)$.

Subcase 3. Suppose that $\hat{\lambda}^*(y) = 1$. Similar to the previous case, the first order condition yields that $\text{E}[\ell'(y - D - \epsilon)D] + \Gamma'(1) - \theta c \mu \geq 0$. As $\Gamma'(\lambda) \geq \hat{\Gamma}'(\lambda)$, we have $\text{E}[\ell'(y - D - \epsilon)D] + \Gamma'(1) - \theta c \mu \geq 0$. Again, applying the first order condition, we have $\lambda^*(y) \geq 1$, i.e., $\lambda^*(y) \geq \hat{\lambda}^*(y)$.

With these properties, we now ready to show that $S \geq \hat{S}$ by considering the following three cases.

Case 1. If $\hat{\lambda}^*(\hat{S}) = 0$, then it is trivial that $\lambda^*(S) \geq \hat{\lambda}^*(\hat{S})$ since $\lambda^*(y) \geq 0$. From (19), we derive the following result: If $\lambda^*(S) \geq 0$, then the first order condition implies that $(\theta - 1) c - \text{E}[\ell'(S - \epsilon)] = 0$
and \((\theta - 1)c - E[\ell'(\hat{S} - \varepsilon)] = 0\), i.e., \(S = \hat{S}\). If \(\lambda^*(S) > 0\), then the first order condition means that 
\[(\theta - 1)c - E[\ell'(S - \lambda^*(S)D - \varepsilon)] = 0.\] 
As \((\theta - 1)c - E[\ell'(\hat{S} - \varepsilon)] = 0\), it is trivial that \(S \geq \hat{S}\) since \(\lambda^*(S) > 0\) and \(-\ell'(x)\) is decreasing in \(x\).

Case 2. Suppose that \(\hat{\lambda}^*(\hat{S}) \in (0, 1)\). As \(\hat{J}_\infty(y, \lambda)\) is jointly concave in \((y, \lambda)\), \(\hat{J}_\infty(y, \hat{\lambda}^*(y))\) is concave in \(y\).

Taking derivative of \(\hat{J}_\infty(y, \hat{\lambda}^*(y))\) with respect to \(y\) and simplifying, we have

\[
\frac{d\hat{J}_\infty(y, \hat{\lambda}^*(y))}{dy} = (\theta - 1)c - E[\ell'(y - \hat{\lambda}^*(y)D - \varepsilon)] + \left\{E[\ell'(y - \hat{\lambda}^*(y)D - \varepsilon)] + \hat{\Gamma}'(\hat{\lambda}^*(y)) - \theta c\mu\right\}\frac{d\hat{\lambda}^*(y)}{dy}.
\]

As the first order condition shows that \(E[\ell'(y - \hat{\lambda}^*(y)D - \varepsilon)] + \hat{\Gamma}'(\hat{\lambda}^*(y)) - \theta c\mu = 0\), we derive

\[
\frac{d\hat{J}_\infty(y, \hat{\lambda}^*(y))}{dy} = (\theta - 1)c - E[\ell'(y - \hat{\lambda}^*(y)D - \varepsilon)]
\]

and

\[
\frac{d\hat{J}_\infty(y, \hat{\lambda}^*(y))}{dy}
\]

\[
\bigg|_{y = \hat{S}} = (\theta - 1)c - E[\ell'(\hat{S} - \hat{\lambda}^*(\hat{S})D - \varepsilon)] = 0.
\]

If \(\lambda^*(S) \in (0, m)\). Similar to the previous analysis, we have

\[
\frac{dJ_\infty(y, \lambda^*(y))}{dy}
\]

\[
\bigg|_{y = \hat{S}} = (\theta - 1)c - E[\ell'(S - \lambda^*(S)D - \varepsilon)] = 0.
\]

Note that both \(\frac{dJ_\infty(y, \hat{\lambda}^*(y))}{dy}\) and \(\frac{dJ_\infty(y, \lambda^*(y))}{dy}\) are decreasing in \(y\), then both \(-E[\ell'(y - \hat{\lambda}^*(y)D - \varepsilon)]\) and \(-E[\ell'(y - \lambda^*(y)D - \varepsilon)]\) are also decreasing in \(y\). As \(\lambda^*(y) \geq \hat{\lambda}^*(y)\), we have \(-E[\ell'(y - \lambda^*(y)D - \varepsilon)] \geq -E[\ell'(y - \hat{\lambda}^*(y)D - \varepsilon)]\), i.e., \(\frac{dJ_\infty(y, \lambda^*(y))}{dy} \geq \frac{dJ_\infty(y, \hat{\lambda}^*(y))}{dy}\). Therefore, expressions (5.35) and (5.36) imply \(S \geq \hat{S}\).

If \(\lambda^*(S) = m\), then the first order condition means that \((\theta - 1)c - E[\ell'(S - mD - \varepsilon)] = 0\). As \((\theta - 1)c - E[\ell'(\hat{S} - \hat{\lambda}^*(\hat{S})D - \varepsilon)] = 0\) and \(\hat{\lambda}^*(\hat{S}) < 1 < m\), we have \(S \geq \hat{S}\).

Case 3. Suppose that \(\hat{\lambda}^*(\hat{S}) = 1\). Similar to the previous case, the first order condition yields that \((\theta - 1)c - E[\ell'(\hat{S} - D - \varepsilon)] = 0\) and \(E[\ell'(\hat{S} - D - \varepsilon)] + \hat{\Gamma}'(1) - \theta c\mu \geq 0\), i.e., \(\hat{\Gamma}'(1) - \theta c\mu \geq -E[\ell'(\hat{S} - D - \varepsilon)]\). As \(\ell'(x)\) is increasing in \(x\), we have \(-E[\ell'(\hat{S} - D - \varepsilon)] \geq -E[\ell'(\hat{S} - D - \varepsilon)]\mu\) and \(\hat{\Gamma}'(1) - \theta c\mu \geq -E[\ell'(\hat{S} - D - \varepsilon)]\mu\). Note that \(\hat{\Gamma}'(1) \leq 0\) and \(\mu > 0\), we obtain \(-E[\ell'(\hat{S} - D - \varepsilon)] < 0\), which is in contradiction to \((\theta - 1)c - E[\ell'(\hat{S} - D - \varepsilon)] = 0\). Therefore, \(\hat{\lambda}^*(\hat{S}) = 1\ need not be discussed in the optimal situation.

Finally, we show that \(\lambda^* \geq \hat{\lambda}^*\) by considering the following three cases.

Case 1. If \(\hat{\lambda}^* = 0\), then it is trivial that \(\lambda^* \geq \hat{\lambda}^*\) since \(\lambda^* \geq 0\).

Case 2. Suppose that \(\lambda^* \in (0, 1)\). The first order condition implies that \((\theta - 1)c - E[\ell'(S - \lambda^*D - \varepsilon)] = 0\) and \((\theta - 1)c - E[\ell'(\hat{S} - \lambda^*D - \varepsilon)] = 0\). Suppose \(\lambda^* < \hat{\lambda}^*\). As \(S \geq \hat{S}\), we have \(S - \lambda^*D - \varepsilon > \hat{S} - \hat{\lambda}^*D - \varepsilon\) with probability one. As \(\ell'(x)\) is increasing in \(x\), we obtain \(E[\ell'(S - \lambda^*D - \varepsilon)] > E[\ell'(\hat{S} - \hat{\lambda}^*D - \varepsilon)]\), which is in contradiction to \((\theta - 1)c - E[\ell'(S - \lambda^*D - \varepsilon)] = 0\) and \((\theta - 1)c - E[\ell'(\hat{S} - \hat{\lambda}^*D - \varepsilon)] = 0\). Therefore, \(\lambda^* \geq \hat{\lambda}^*\).

Case 3. Suppose that \(\lambda^* = 1\). Note that \(\hat{\lambda}^* = \hat{\lambda}(\hat{S})\) and \(\hat{\lambda}(\hat{S}) = 1\ need not be considered in the optimal situation. As a result, \(\lambda^* = 1\ need not be considered.

\(\square\)

**Proof of Proposition 6.** As \(\Delta = \frac{1}{\theta^2} \left[J_\infty(S, \lambda^*) - J_\infty(\hat{S}, \hat{\lambda}^*)\right]\), we just prove that \(J_\infty(S, \lambda^*) - J_\infty(\hat{S}, \hat{\lambda}^*)\)
is decreasing in $c$. Taking derivative of $J_\infty(S, \lambda^o)$ and $\hat{J}_\infty(S, \hat{\lambda}^o)$ with respect to $c$ and simplifying, we have

\[
\frac{dJ_\infty(S, \lambda^o)}{dc} = (\theta - 1)S - \theta \mu \lambda^o + \frac{\partial J_\infty(S, \lambda^o)}{\partial y} \cdot \frac{dS}{dc} + \frac{\partial J_\infty(S, \lambda^o)}{\partial \lambda} \cdot \frac{\lambda^o}{dc},
\]

(5.37)

\[
\frac{d\hat{J}_\infty(S, \hat{\lambda}^o)}{dc} = (\theta - 1)\hat{S} - \theta \mu \hat{\lambda}^o + \frac{\partial \hat{J}_\infty(S, \hat{\lambda}^o)}{\partial y} \cdot \frac{d\hat{S}}{dc} + \frac{\partial \hat{J}_\infty(S, \hat{\lambda}^o)}{\partial \lambda} \cdot \frac{\hat{\lambda}^o}{dc}.
\]

(5.38)

As $S$ and $\hat{S}$ are interior points, the first order condition means that $\frac{\partial J_\infty(S, \lambda^o)}{\partial y} = \frac{\partial \hat{J}_\infty(S, \hat{\lambda}^o)}{\partial y} = 0$. From the proof of Proposition 5, we have $\hat{\lambda}^o < 1$ since $\lambda^o = 1$ need not be considered. As $\hat{\lambda}^o = 0$ is a trivial case, we just consider the per-unit ordering cost $c$ such that $\hat{\lambda}^o > 0$. Therefore, $\frac{\partial J_\infty(S, \lambda^o)}{\partial \lambda} = 0$. Similarly, $\frac{\partial \hat{J}_\infty(S, \hat{\lambda}^o)}{\partial \lambda} = 0$. (5.37) and (5.38) imply

\[
\frac{dJ_\infty(S, \lambda^o)}{dc} - \frac{d\hat{J}_\infty(S, \hat{\lambda}^o)}{dc} = (\theta - 1)(S - \hat{S}) - \theta \mu (\lambda^o - \hat{\lambda}^o).
\]

Note that $S \geq \hat{S}$ and $\lambda^o \geq \hat{\lambda}^o$, then $\Delta$ is decreasing in $c$, which completes the proof. \hfill $\square$  

**Proof of Proposition 7.** Note that $\hat{V}(0)$ is independent of $a$, we next prove that $V(0)$ is increasing in $a$, i.e., $J_\infty(S, \lambda^o)$ is increasing in $a$. Applying the definitions of $S$ and $\lambda^o$ in (19), $\Gamma(\lambda)$ in (5.31) and $\hat{\Gamma}(\lambda)$ in (18), we have

\[
J_\infty(S, \lambda^o) = \max_{y, \lambda \in [0, m]} \left\{ (\theta - 1)cy - E[\ell(y - \lambda D - \varepsilon)] - \theta c \mu \lambda \right. \\
+ \left. \max_{0 < \frac{\lambda}{m} \leq \frac{\lambda^o}{m} \leq \frac{\hat{\lambda}^o}{m}} \{ \hat{\Gamma}(\lambda - (m - 1)\lambda^o) + (a - 1)\hat{\Gamma}(\lambda^o) \} \right\}.
\]

Note that $\hat{\Gamma}(\lambda)$ is nonnegative for any $\lambda$. It follows that $J_\infty(S, \lambda^o)$ is increasing in $a$. \hfill $\square$

**Example 5.1 (Subsection 3.1).** Consider a single period problem in which the market share functions are defined by the functions in (5). Suppose that the value of one unit product $V$ follows a distribution in $[0, 1]$ with a cumulative distribution function whose inverse is $G^{-1}(x) = 3x^3 - 3.2x^2 + 1.2x$. As $xG^{-1}(1 - x) = -3x^4 + 5.8x^3 - 3.8x^2 + x$ is not concave, the proof of Proposition 1 implies that $V$ does not have an increasing virtual value. And hence Assumptions 1, 2 and 3 are all violated. It can be shown that when the inventory increases, the unit-sales mode follows the pattern of quoted $\rightarrow$ removed $\rightarrow$ quoted $\rightarrow$ removed. Hence, the optimal pricing strategy does not have to follow the strategy shown in Figure 1 if $V$ does not have an increasing virtual value.

**Example 5.2 (Discussion Below Proposition 3).** We consider that $V$ follows a distribution in $[0, 1]$ with $G^{-1}(x) = 3x^3 - 5x^2 + 3x$. It is straightforward to verify that $xG^{-1}(1 - x) = -3x^4 + 4x^3 - 2x^2 + x$ is concave, and hence by Proposition 1 and its proof, Assumptions 1, 2 and 3 are all satisfied.

Figure 3 illustrates why a price surcharge is quoted when $\lambda_1 = 0.6$. When both prices $p_{11}$ and $p_{22}$ are quoted, the expected revenue in period $t$ is $p_{1t}(\lambda_{1t}, \frac{\lambda_{1t} - \lambda_{1u}}{m}) \times \lambda_{1t} + p_{2t}(\lambda_{1t}, \frac{\lambda_{1t} - \lambda_{1u}}{m}) / m \times (\lambda_1 - \lambda_{1t})$. The area of the shaded rectangles in Figure 3 corresponds to the revenue when $\lambda_{1t} = 0.21$, which also corresponds to the maximum revenue given $\lambda_1 = 0.6$. As shown in Figure 3, the functions $p_{1t}(\lambda_{1t}, \frac{\lambda_{1t} - \lambda_{1u}}{m})$ and $p_{2t}(\lambda_{1t}, \frac{\lambda_{1t} - \lambda_{1u}}{m}) / m$ intersect when $\lambda_{1t}$ lies around 0.1 and the function $p_{2t}(\lambda_{1t}, \frac{\lambda_{1t} - \lambda_{1u}}{m}) / m$ is increasing in $\lambda_{1t}$. Therefore, the expected revenue is maximized when $p_{2t}(\lambda_{1t}, \frac{\lambda_{1t} - \lambda_{1u}}{m}) / m > p_{1t}(\lambda_{1t}, \frac{\lambda_{1t} - \lambda_{1u}}{m})$, i.e., a surcharge is quoted for quantity sales.
Figure 3: Prices $p_{1t}$ and $p_{2t}$ as Functions of $\lambda_1t$ When $\lambda_t = 0.6$ for Example 5.2

B Other Models

B.1 The Model with a Fixed Ordering Cost

For the situation in which there is a fixed ordering cost in each period, Chen and Simchi-Levi (2004a) studied the model with time-based price differentiation and showed that the optimal inventory control policy generally does not have an $(s, S)$ structure. Instead, it follows an $(s, S, A)$ structure. It is fairly straightforward to show that the optimal inventory control policy for our model also takes the $(s, S, A)$ structure. As far as the current framework is concerned, such a policy structure is the best analytical result that we can expect for our problem because it involves quantity-based price differentiation, which is more complex than the problem addressed in Chen and Simchi-Levi (2004a). For the selling/pricing decision, examples can be constructed to show that the pattern exhibited in Figure 1 no longer holds even in the presence of Assumptions 1-3. Please see the following example.

Example 5.3. We consider a two-period problem with $\theta = 1$, $c = \frac{33}{64}$, $L(x) = 0.0372x^+ + x^-$, $D = 1$, $\varepsilon = 0$, $G(x) = x$ for $x \in [0, 1]$ (the utility model is used), $m = 2$, $K = \frac{1}{128}$, and $a = 1.5$. The end of horizon value function $V_3(x)$ is set to $V_3(x) = 0.0372x^+ - \frac{33}{64}x^-$. It can be shown that in both first and second periods, the optimal inventory policy follows an $(s, S)$ policy with $s = 0.1538$ and $S = 0.2422$. In the first period, quantity-sales mode appears on two disjoint regions $y^*_1 \in [0, 0.25)$ and $y^*_1 \in [0.3798, +\infty)$, which destroys the structure of optimal pricing policy demonstrated in Figure 1 for the model without a fixed ordering cost.

This difference comes from the fact that $\lambda^*_t(y)$ is an increasing function of $y$ in the model without a fixed ordering cost (see Lemma 1) while this property does not hold when there is a fixed ordering cost. In the model without a fixed ordering cost, Lemma 5.1 shows us that the concavity of $W_{t+1}(x)$ implies the supermodularity of $J_t(y, \lambda)$. However, in the model with a fixed ordering cost, $W^K_{t+1}(x)$ is not concave and hence $J^K_t(y, \lambda)$ is not supermodular in $(y, \lambda)$ any more, which destroys the monotonicity of $\lambda^*_t(y)$. The intuition behind Example 5.3 can be explained as follows. If $y^*_1 \in [0.25, 0.3798]$, then the firm should use quantity sales to increase demand so that it will most likely order in the next period; if $y^*_1 \in [0.3798, 0.4814]$, then the firm would not use quantity sales to push up demand because the leftover inventory in period 1 can satisfy future demand such that no
order is placed in the next period and hence a fixed ordering cost can be saved; if \( y^*_t \geq 0.4814 \), the firm must use quantity sales mode to clear up the inventory.

For the infinite horizon case, we can show that the optimal inventory control policy is an \((s, S)\) policy by using a similar methodology in Chen and Simchi-Levi (2004b). However, once again, a counter example similar to Example 5.3 can be constructed to show that the optimal selling/ pricing pattern does not follow Figure 1, even in the infinite-horizon stationary setting.

### B.2 Positive Leadtime

In this section, we consider the joint inventory and pricing problem with a positive constant leadtime. For a multiplicative demand model, it is well-known that there is no nice structure even if the firm does not use quantity-based price differentiation (see Pang et al. 2012). For our model with quantity-based price differentiation, the optimal policy illustrated by Figure 1 does not exist because the property like Lemma 1 (a) does not hold under a positive leadtime. Fortunately, the structure of optimal policy can somehow be preserved for the additive demand model, i.e., \( D_t \) is deterministic and hence \( \Pr(D_t = \mu_t) = 1 \).

Denote \( L \geq 1 \) as the leadtime and \( q_t \) as the order quantity at period \( t \). The system state is represented by the \( L \)-vector \( \mathbf{x}_t = (x_{0,t}, x_{1,t}, \ldots, x_{L-1,t}) \), where \( x_{0,t} \) denotes the inventory level at the beginning of period \( t \) after the order due in period \( t \) arrives and \( x_{i,t} (i = 1, 2, \ldots, L-1) \) denotes the order quantity in period \( t+i-L \), i.e., \( x_{1,t} = q_{t+1-L}, \ldots, x_{L-1,t} = q_{t-1} \). Let \( \lambda_t = \lambda_{1t} + m\lambda_{2t} \). The dynamics of the system can be expressed as \( \mathbf{x}_{t+1} = (x_{0,t} - \lambda_t \mu - \varepsilon_t + x_{1,t}, x_{2,t}, \ldots, x_{L-1,t}, q_t) \). Let \( v_{j,t} = \sum_{i=0}^{j} x_{i,t} \) for \( j = 0, 1, \ldots, L-1 \) and denote \( \mathbf{v}_t = (v_{0,t}, v_{1,t}, \ldots, v_{L-1,t}) \). The optimality equation can be written as

\[
V_t(v_{0,t}, v_{1,t}, \ldots, v_{L-1,t}) = \max_{y \geq v_{L-1,t}, 0 \leq \lambda_t \leq m} \mathbb{E}\left\{ -c(y - v_{L-1,t}) - \ell(v_{0,t} - \lambda_t \mu - \varepsilon_t) + \Gamma(\lambda_t) + \theta V_{t+1}(v_{1,t} - \lambda_t \mu - \varepsilon_t, v_{L-1,t} - \lambda_t \mu - \varepsilon_t, y - \lambda_t \mu - \varepsilon_t) \right\},
\]

where \( \Gamma(\lambda_t) \) is defined in (12) and Assumptions 1, 2 and 3 are satisfied. Let \( (y_t(\mathbf{v}_t), \lambda_t(\mathbf{v}_t)) \) denote the least optimal solution of dynamic program (5.39). Hence the optimal order quantity \( q_t(\mathbf{v}_t) \) follows that \( q_t(\mathbf{v}_t) = y_t(\mathbf{v}_t) - v_{L-1,t} \) if \( y_t(\mathbf{v}_t) > v_{L-1,t} \), and \( q_t(\mathbf{v}_t) = 0 \) if \( y_t(\mathbf{v}_t) \leq v_{L-1,t} \).

Following Theorem 1 in Pang et al. (2012), we can show that \( V_t(v_{0,t}, v_{1,t}, \ldots, v_{L-1,t}) \) is \( L^\# \)-concave, which implies that \( \lambda_t(\mathbf{v}_t) \) is nondecreasing in \( \mathbf{v}_t \). By this monotonicity, we can define

\[
\underline{v}_{L-1,t}(v_{0,t}, \ldots, v_{L-2,t}) = \sup \{ v_{L-1,t} : \lambda_t(v_{0,t}, \ldots, v_{L-2,t}, v_{L-1,t}) \leq \lambda \text{ for any } v_{0,t}, \ldots, v_{L-2,t} \},
\]

\[
\overline{v}_{L-1,t}(v_{0,t}, \ldots, v_{L-2,t}) = \inf \{ v_{L-1,t} : \lambda_t(v_{0,t}, \ldots, v_{L-2,t}, v_{L-1,t}) > \lambda \text{ for any } v_{0,t}, \ldots, v_{L-2,t} \},
\]

where \( \lambda \) is the threshold on \( \lambda \) below which it is optimal to use unit sales only and \( \lambda \) is the threshold on \( \lambda \) above which it is optimal to use quantity sales only. As \( \lambda \leq \overline{\lambda} \), the monotonicity of \( \lambda_t(\mathbf{v}_t) \) implies that \( \underline{v}_{L-1,t}(v_{0,t}, \ldots, v_{L-2,t}) \leq \overline{v}_{L-1,t}(v_{0,t}, \ldots, v_{L-2,t}) \) and hence the following proposition holds.

**Proposition 5.3.** For \( t = 1, \ldots, n \) and any \( \mathbf{v}_t = (v_{0,t}, v_{1,t}, \ldots, v_{L-1,t}) \), we have the following optimal structure.

(a) If \( v_{L-1,t} \leq \underline{v}_{L-1,t}(v_{0,t}, \ldots, v_{L-2,t}) \), then the firm sets a uniform price only, and the corresponding unit-sales market share is \( \lambda_t(\mathbf{v}_t) \).
we have 
\[ \lambda_1(v_i) = m \lambda_2^*(\lambda_1(v_i)) \] 
and 
\[ \lambda_2^*(\lambda_1(v_i)) \]
respectively.

(c) If \( v_{L-1,t} > \bar{v}_{L-1,t} \), then the firm uses quantity-sales mode only, and the corresponding market share is \( \lambda_2^*(v_i) = \lambda_1(v_i)/m \).

Although the results in Section 2 somehow hold for the model with a positive leadtime, those results in Section 3 are not applicable here because they depend heavily on the feature of a single state while the model with a positive leadtime is of multiple states.

**B.3 Multiple Sales Modes**

In this section, we consider a valuation-based demand model with \( I \) sales modes, i.e., the firm quotes price \( p_i \) for the \( i \)th sales mode, which consists of \( m_i \) units with \( m_1 < m_2 < \ldots < m_I \). We denote \( a_iV \) as the value of the \( i \)th mode with \( a_1 \leq a_2 \leq \ldots \leq a_I \). Then the utility of purchasing \( i \)th mode is \( a_iV - p_i \). Let \( \mathbf{p} = (p_1, \ldots, p_I) \) and \( \lambda = (\lambda_1, \ldots, \lambda_I) \).

The market share of the \( i \)th sales mode is given by

\[
\lambda_i(\mathbf{p}) = \Pr(a_iV - p_i \geq \max_j \{a_jV - p_j \}, a_iV - p_i \geq 0) = \Pr \left( \max_j \left\{ \frac{p_j - p_i}{a_i - a_j} \right\} \leq V \leq \min_{i>1} \left\{ \frac{p_i - p_i}{a_i - a_i} \right\} \right) \geq \frac{p_i}{a_i} \right).
\]

For \( i = 1 \), Equation (5.40) implies that it is sufficient to consider the prices \( p_1 \) and \( p_2 \) satisfying \( \frac{p_1}{a_1} \leq \frac{p_2}{a_2 - a_1} \). Otherwise, the probability of choosing the first mode would be zero. Similarly, for \( i = 2 \), Equation (5.40) implies that it is sufficient to consider the prices \( p_1, p_2 \) and \( p_3 \) satisfying \( \frac{p_2 - p_1}{a_2 - a_1} \leq \frac{p_3 - p_2}{a_3 - a_2} \). Following a similar argument, the \( i \)th sales mode implies that we only need to consider the prices that satisfy \( \frac{p_i - p_{i-1}}{a_i - a_{i-1}} \leq \frac{p_{i+1} - p_i}{a_{i+1} - a_i} \).

In summary, we only need to consider the set

\[
\mathcal{P} = \left\{ \mathbf{p} : \frac{p_1}{a_1} \leq \frac{p_2 - p_1}{a_2 - a_1} \leq \ldots \leq \frac{p_i - p_{i-1}}{a_i - a_{i-1}} \leq \ldots \leq \frac{p_I - p_{I-1}}{a_I - a_{I-1}} \right\}.
\]

For any \( \mathbf{p} \in \mathcal{P} \), we can show that \( \min_{i>1} \left\{ \frac{p_i - p_1}{a_i - a_1} \right\} = \frac{p_i - p_{i-1}}{a_i - a_{i-1}} \) and \( \max_{j<i} \left\{ \frac{p_i - p_j}{a_i - a_j} \right\} = \frac{p_i - p_{i-1}}{a_i - a_{i-1}} \geq \frac{p_i}{a_i} \). Therefore, we have \( \lambda_i(\mathbf{p}) = \Pr \left( \frac{p_i - p_{i-1}}{a_i - a_{i-1}} \leq V \leq \frac{p_{i+1} - p_i}{a_{i+1} - a_i}, V \geq \frac{p_i}{a_i} \right) = G \left( \frac{p_i - p_{i-1}}{a_i - a_{i-1}} \right) - G \left( \frac{p_{i+1} - p_i}{a_{i+1} - a_i} \right), i = 1, \ldots, I \), where \( G(x) \) is the distribution of \( V \) and we assume that \( p_0 = 0, a_0 = 0 \) and \( p_{I+1} = +\infty \). By solving these expressions, we have \( p_1(\lambda) = a_1G^{-1}(1 - \sum_{j=1}^I \lambda_j) \) and

\[
p_i(\lambda) = a_iG^{-1} \left( 1 - \sum_{j=1}^I \lambda_j \right) + (a_2 - a_1)G^{-1} \left( 1 - \sum_{j=2}^I \lambda_j \right) + \ldots + (a_i - a_{i-1})G^{-1} \left( 1 - \sum_{j=i}^I \lambda_j \right), \quad 2 \leq i \leq I.
\]

Hence, the revenue function

\[
R(\lambda_1, \lambda_2, \ldots, \lambda_I) = \sum_{i=1}^I p_i(\lambda)\lambda_i = \left[ a_1 \zeta \left( \sum_{i=1}^I \lambda_i \right) + (a_2 - a_1) \zeta \left( \sum_{i=2}^I \lambda_i \right) \right] + (a_3 - a_2) \zeta \left( \sum_{i=3}^I \lambda_i \right) + \ldots + (a_I - a_{I-1}) \zeta(\lambda_I),
\]

38
where $\zeta(x) = xG^{-1}(1 - x)$. If $V$ has an increasing virtual value function, $\zeta(x)$ is a concave function, which implies that $R(\lambda_1, \lambda_2, ..., \lambda_I)$ is jointly concave in $(\lambda_1, \lambda_2, ..., \lambda_I)$. Therefore, Theorem 1 holds, i.e., the optimal inventory policy still follows a base-stock type. Define $\lambda = \sum_{i=1}^I m_i \lambda_i$ that determine the total sales, we still have the property that optimal $\lambda^*(y)$ is an increasing function of order-up-to level $y$ as shown in Lemma 1.

The structure of the optimal pricing policy derived in Section 2 relies on the fact that for each sales mode, there is a critical inventory level for the sales mode to disappear/appear (see Theorem 2). To see whether this property still holds, we let $\lambda_1 = (\lambda - \sum_{i=2}^I m_i \lambda_i)/m_1$ and write the revenue function as

$$
\hat{R}(\lambda, \lambda_2, ..., \lambda_I) = \left[ a_1 \zeta\left(\sum_{i=2}^I \lambda_i + \frac{\lambda - \sum_{i=2}^I m_i \lambda_i}{m_1}\right) + (a_2 - a_1) \zeta\left(\sum_{i=2}^I \lambda_i\right) + (a_3 - a_2) \zeta\left(\sum_{i=3}^I \lambda_i\right) +
\right.
\left.
\vdots
\right.
\left.
+ (a_I - a_{I-1}) \varphi(\lambda_I)\right].
$$

For any $\lambda$, let $\lambda^*_i(\lambda)$, $i = 2, \ldots, I$ be the optimal solution that maximizes $\hat{R}(\lambda, \lambda_2, ..., \lambda_I)$. If $V$ follows the uniform distribution, then $\zeta(x)$ is a quadratic function and hence $\lambda^*_i(\lambda)$ is a linear function of $\lambda$. As optimal $\lambda^*(y)$ is an increasing function of order-up-to level $y$, there is a critical inventory level for each sales mode to disappear/appear. However, for general distribution of $V$, $\lambda^*_i(\lambda)$ can be a very complicated function of $\lambda$ even if $V$ has an increasing virtual value function (the only assumption made in Section 3). Therefore, the property that there is a critical inventory level for each sales mode to disappear/appear does not necessarily hold (which can easily be demonstrated by counter-examples).