Dynamic Competitive Newsvendors with Service-Sensitive Demands

Liming LIU
Weixin SHANG
Shaohua WU

Available at: https://works.bepress.com/limingliu/21/
Dynamic Competitive Newsvendors with Service-Sensitive Demands

Liming Liu  
Department of Logistics, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong,  
lgtliu@polyu.edu.hk

Weixin Shang  
Department of Information and Systems Management, Hong Kong University of Science and Technology,  
Clear Water Bay, Kowloon, Hong Kong, shangwx@ust.hk

Shaohua Wu  
China Merchants Holdings (International) Company Limited, Shun Tak Centre, Central, Hong Kong,  
shwu@cmhi.com.hk

When two firms compete for service-sensitive demands based on their product availability, their actions will affect the future market share reallocation. This problem was first studied by Hall and Porteus (2000) using a dynamic game model. We extend their work by incorporating a general demand model, which enables us to obtain properties that reveal the dynamics of the game and the behavior of the players. In particular, we provide conditions under which the market share of a firm has a positive value and give it an upper bound. We further extend the game competition model to an infinite-horizon setting. We prove that there exists a stationary equilibrium policy and that the dynamic equilibrium policy always converges to a stationary equilibrium policy. We demonstrate that demand patterns will dictate how firms compete rationally and show the likely outcomes of the competition.

Key words: service-sensitive demand; availability competition; demand model; dynamic game; feedback Nash equilibrium; stationary policy; order quantity structure

History: Received: June 18, 2004; accepted: October 12, 2005.

1. Introduction

This paper studies competition in product availability between two firms. In a nonmonopoly market, the immediate customer reaction to a stockout is either accepting the delayed delivery (backorder, often with some incentive) or taking the order elsewhere (lost sales). The standard inventory control theory handles the immediate impact of a stockout satisfactorily from the supplier’s point of view. When the second action is taken, however, the supplier suffers not only a loss of sales but also a loss of goodwill, which will likely affect future demand for the product (Schwartz 1966). Clearly, two factors should be considered when suppliers formulate their production and supply strategies in a competitive market: the impact of a strategy on future demands and the possible actions and reactions of the competitors. Issues related to these two factors have been studied extensively in the literature, roughly in two research streams: that considering the impact of a service strategy on future demands and that considering product substitution and competition.

Schwartz (1966) pointed out that backorder costs for stockouts cannot fully describe the impact of loss of goodwill in the future. He proposed a “perturbed demand” model, in which the demand in any period equals the potential demand multiplied by a factor that depends on observed customer fulfillment rate. This research has been followed and extended in a number of works, including Schwartz (1970), Hill (1976), Caine and Plaut (1976), and Robinson (1990), among others.

The literature on product substitution deals with situations in which two or more suppliers with substitutable products compete in the same market. Examples include, but are not limited to, McGillivray and Silver (1978), Parlar and Goyal (1984), Parlar (1985, 1988), Li et al. (1990), Lau (1991), and Wang and Parlar.
The focus here is on developing strategies for competition in the current period. Future effects are ignored.

Similarly, a number of authors have developed strategies for competition based on product availability in a single period. Li (1992) considered competition in production speed in a buyer’s market, assuming that a demand will be filled by the supplier that produces the next available product first. This line of research has been followed and extended by Ernst and Cohen (1992), Li and Lee (1994), and Lederer and Li (1997).

Lippman and McCardle (1997) introduced competition into the standard newsvendor problem. In their model, two firms make ordering decisions at the beginning of a period to compete for the demand in the current period. When a shortage happens at one firm, the unmet demand switches to the other firm. Along the same line, Netessine et al. (2006) considered a two-firm competition problem in a reorder point system setting. When a stockout occurs at one firm, the unmet demand will either be backordered or will switch to a competitor immediately. Stationary optimal ordering strategies are developed under four different scenarios. Because future demand is not affected by current activities, the problem is essentially a one-period problem. Bernstein and Federgruen (2004) considered price and service-sensitive demands in a one-period setting, using a multiplicative demand model. They showed that the equilibrium in an infinite-horizon setting is the same as in the one-period setting.

From our brief literature review one can see that, since Schwartz’s work (1966), the two factors—competition in product availability and its future effect—have more or less been studied separately. The only exception is Hall and Porteus (2000), who considered service competition among multiple firms in a dynamic setting in which product availability in the current period affects the expected demand in the next period.

In this paper, we extend the work by Hall and Porteus (2000) by incorporating a general demand model and relax their key assumption on the service failure pattern. The general demand model enables us to carry out more in-depth analysis of firm behavior under competition. For example, it can be shown that different demand patterns lead to significantly different valuations by a firm on its market share. We prove that the optimal order quantity for a firm under availability competition is determined by a modified newsvendor critical fractile, and we give explicit formulas for calculating the unique feedback Nash equilibrium. We further extend this competition model to an infinite planning horizon and show that the dynamic equilibrium ordering policy always converges to a stationary equilibrium policy. We demonstrate that the underlying demand pattern plays a key role in determining whether a firm can survive the competition in the long run.

This paper is organized as follows. In §2, we introduce a general demand model and discuss its properties. We then present the dynamic product availability competition model. In §3, we show the existence of and provide explicit formulas for the unique feedback Nash equilibrium. In §4, we analyze the behavior of the firms under competition, especially, how the value of market share is affected by the demand pattern. In §5, we discuss availability competition in an infinite planning horizon. We derive the stationary equilibrium policy and show its relationship with the dynamic equilibrium policy. We further analyze whether and how a competing firm can survive under such competition.

2. Model Basics
Consider the following scenario: Two firms selling homogeneous or substitutable products compete over $T$ periods in the same market. Competition arises because unsatisfied demands in one period will cause the loss of new demands in the next period proportionally, forcing the firms to strategize their ordering decisions to maintain appropriate product-availability levels. In this section, we first introduce our demand model and then formulate the problem as a dynamic game.

2.1. Demand Model
Our basic assumptions for the demand are as follows. Demands at each firm can be nonstationary over the finite planning horizon, but they follow some pattern characterized by a density function and some parameters. Let $\epsilon$ be a continuous random variable with mean 0 and density function $f_\epsilon(x)$, which is continuous on the support $S = \{x: f_\epsilon(x) > 0\}$ and is referred to
as the seed density. We define, for some nonnegative constants \(a, b, \) and \(\mu > 0\), a random demand \(D\) by

\[
D = \mu + (a\mu + b)\epsilon. \tag{1}
\]

In other words, this demand consists of a deterministic mean \(\mu\) and a random noise. The spread of the random noise is an affine function of the mean demand. When \(f_0(x)\), \(a\), and \(b\) are given, we say that the demand pattern is known. For a given demand pattern, the demand distribution is uniquely determined by the mean \(\mu\).

Such a demand model includes many widely used demand models as special cases. We may construct different demands by selecting different seed densities and constant parameters. We will focus on the following two demand patterns:

**Additive demand model.** If we set \(a = 0, b > 0\), then 

\[
D = \mu + \epsilon \quad \text{with} \quad \epsilon = b\epsilon \quad \text{and} \quad E[\epsilon'] = 0.
\]

**Multiplicative demand model.** If we set \(a > 0\) and \(b = 0\), then 

\[
D = \mu\epsilon \quad \text{with} \quad \epsilon = a\epsilon + 1 \quad \text{and} \quad E[\epsilon'] = 1.
\]

In a market place, a firm may have a well-established brand (or provide superior after sales services) and reputation and thus enjoy one cluster of customers who will always buy from this firm. These customers form a demand base for this firm. In addition, this firm also receives random demands from other customers. The overall demand for this firm can be modeled as additive demand. If a firm does not have such a well-established brand and reputation, it will not have the luxury of a fixed demand base, and the multiplicative demand model defined above may be a better fit.

A key assumption in Hall and Porteus’ work is that the expected unmet demand equals \(\mu H(q/\mu)\) for some twice differentiable, decreasing, and strictly convex function \(H\), when the order quantity is \(q\). Essentially, they assume a multiplicative demand model. The model defined in (1) only requires the continuity of the seed density \(f_0\) on its support. As we will see, additive and multiplicative demands lead to very different valuations of market shares and competition outcomes, which shows the importance of a general demand framework to the analysis of a dynamic competitive market. We now provide some general properties of the demand model (1).

Let us assume that the nonnegative constant \(a\) is sufficiently small so that the support of the seed density is on the right side of \(-1/a\), that is, \(S \subset \{x: x > -1/a\}\). For the multiplicative demand model, this assumption guarantees that the realized demand is always positive. Let \(f(x)\) denote the density of \(D\) and \(F_0\) and \(F\) the distribution functions corresponding to \(f_0\) and \(f\), respectively. Let \(s_0(\beta) = F_0^{-1}(\beta)\) and \(s(\beta) = F^{-1}(\beta)\) be the levels of inventory required to satisfy all demands with a probability \(\beta\) \((0 < \beta < 1)\). We obtain the following interesting and important properties.

**Lemma 1.** For the demand defined by Equation (1), we have for any \(\beta \in (0, 1)\)

\[
s(\beta) = [as_0(\beta) + 1]\mu + bs_0(\beta), \tag{2}
\]

\[
\int_0^{s(\beta)} xf(x) dx = [ay_0(\beta) + b]\mu + by_0(\beta), \tag{3}
\]

where \(y_0(\beta) = \int_{-1/\alpha}^{s(\beta)} x f_0(x) dx\). Furthermore, define

\[
k_0(\beta) = -(1 - \beta)s_0(\beta) - y_0(\beta), \tag{4}
\]

then \(k_0(\beta) \in [0, 1/\alpha]\) and \(k_0(\beta) = -(1 - \beta)/f_0(s_0(\beta)) < 0\).

The proof of this lemma, and all the other proofs, are in the appendix.

Given that the demand is \(\epsilon\), Equation (4) shows that \(-k_0(\beta)\) gives the expected satisfied demand when the order quantity is \(s_0(\beta)\). It is also easy to verify that, for demand \(D\), the expected unsatisfied demand is

\[
E[(D - q)^+] = ak_0(F_0\left[q - \frac{\mu}{a\mu + b}\right])\mu + bk_0(F_0\left[q - \frac{\mu}{a\mu + b}\right]), \tag{5}
\]

when the order quantity is \(q\). When demand is multiplicative, or \(b = 0\), the second term on the right-hand side of (5) vanishes, reducing to the model used in Hall and Porteus (2000).

To define the demand patterns for the two firms over \(T\) periods, we use subscripts \(i\) and \(j\) to indicate (different) firms and \(t\) to indicate a time period whenever necessary. For firm \(i\), the demand in period \(t\) is

\[
D_{it} = \mu_i + (a\mu_i + b)\epsilon_{it},
\]

where \(\epsilon_{it}\) is independent and identically distributed across all periods, following the same distribution as that of \(\epsilon\).
2.2. Dynamic Competitive Newsvendors

We assume that two firms (firm 1 and firm 2) provide a perishable homogeneous product or service to a market that is indifferent to the price or quality of the product. In each period, the sequence of events is as follows: At the beginning of a period, each firm observes its market share and makes an ordering decision; then orders arrive at the two firms; realized demands are either satisfied or lost when stock-out occurs; finally, unused products are salvaged and revenue collected. The demand patterns are common knowledge to both firms. The initial market allocation at the beginning of period 1 is also known to both firms. The mechanism that determines how the market share \( \mu_{i} \) is reallocated according to the availability competition in each period is essentially the same as the one used in Hall and Porteus (2000). Thus, its discussion below will be brief.

Customers who encounter stockouts are called disappointed customers. A fixed proportion of disappointed customers will switch from their current supplier to the competitor in the subsequent period. Let \( q_{it} \) be the order quantity and \( U_{it} \) be firm \( i \)'s expected lost sales in period \( t, i = 1, 2; t = 1, 2, \ldots, T \). We have \( U_{it} = E[(D_{it} - q_{it})^{+}] \). The following equation completely characterizes the market share reallocation mechanism:

\[
\mu_{i,t+1} = \mu_{it} - \alpha_{i}U_{it} + \alpha_{j}U_{jt}, \quad \text{for } i = 1, 2; j \neq i,
\]

where \( \alpha_{i}, \alpha_{j} \in (0, 1) \) are the switching probabilities for firms \( i \) and \( j \), respectively. We point out that while the total market size as indicated by (6) is a constant \( =\mu_{0} = \mu_{11} + \mu_{21} \), the general results in §§3 and 4 can be carried when the market size is time dependent. We also point out that while in reality it is possible that some disappointed customers may simply withdraw from this market, such a scenario is not considered in this paper. We further assume that the two firms will remain in this market for the entire \( T \) periods. The reallocation formula (6) can be rewritten as follows:

\[
\mu_{i,t+1} = g_{i}(q_{it}, q_{jt}, \mu_{it}, \mu_{jt})
\]

\[
= \mu_{it} - \alpha_{i} \int_{q_{it}}^{+\infty} (x-q_{it})f_{it}(x)dx + \alpha_{j} \int_{q_{jt}}^{+\infty} (x-q_{jt})f_{jt}(x)dx.
\]

Equation (6) or (7) provides the link between the current ordering decisions and the market reallocation in the next period.

Let \( Q_{i} \equiv \{q_{it}\} \geq 0 \) denote the set of decisions of firm \( i, i = 1, 2 \). The total expected profit for firm \( i, i = 1, 2 \), over the entire planning horizon is

\[
\pi_{i}(Q_{1}, Q_{2}) = \sum_{t=1}^{T} \gamma^{t-1} h_{it}(q_{it}, \mu_{it}) + \gamma^{T} h_{i,T+1}(\mu_{i,T+1}), \quad \text{(8)}
\]

where

\[
h_{it}(q_{it}, \mu_{it}) = r_{i}q_{it} - (r_{i} + c_{i})E[(q_{it} - D_{it})^{+}],
\]

\[
t = 1, \ldots, T \quad \text{(9)}
\]

\[
h_{i,T+1}(\mu_{i,T+1}) = A_{i,T+1} \mu_{i,T+1} + B_{i,T+1}. \quad \text{(10)}
\]

In the above equations, \( r_{i} \) is firm \( i \)'s unit profit or underage cost; \( c_{i} \) is firm \( i \)'s overage cost, that is, the cost of discarding one unit of unsold product at the end of a period; \( \gamma \) is a discount factor with \( 0 < \gamma < 1 \); \( h_{it}(q_{it}, \mu_{it}) \) is the expected profit for firm \( i \) in period \( t \) \( (t = 1, \ldots, T) \), given its market share and its ordering decision; and the last equation defines, through the nonnegative parameters \( A_{i,T+1} \) and \( B_{i,T+1} \), that the terminal value \( h_{i,T+1}(\mu_{i,T+1}) \) is nonnegative and affine in the exit market share \( \mu_{i,T+1} \).

From (8) and (9), we observe that the total expected profit for either firm depends on the replenishment strategy of the competitor, as well as the firm’s own strategy. Assuming that the two firms make their ordering decisions simultaneously at the beginning of each period, the decision processes of the two firms over the planning horizon constitute a dynamic Nash game:

\[
\max_{Q_{i}} \pi_{i}(Q_{1}, Q_{2})
\]

s.t. \( \mu_{i,t+1} = \mu_{it} - \alpha_{i}U_{it} + \alpha_{j}U_{jt}, \)

\[
\text{for } i = 1, 2; j \neq i \text{ and } t = 1, 2, \ldots, T. \quad \text{(11)}
\]

3. Feedback Nash Equilibrium

To analyze the optimization problem (11), we need the concept of a feedback game. Starting from period \( T \), after observing the market share allocation, each firm solves a one-period newsvendor-type problem. Moving to period \( T-1 \), the current ordering decisions will
affect both the profits in period \( T - 1 \) and the market allocation in period \( T \). The single period profit and market reallocation can be quantified using Equations (7) and (9), respectively. With the market reallocation and the optimal ordering decisions in period \( T \), the two firms have full information to determine their best decisions in period \( T - 1 \).

In this way, the two-period game in period \( T - 1 \) can be analyzed. We see that to study a \( t \)-period problem \( (t > 1) \), we have to study a \((t - 1)\)-period problem first. Thus, for the \( T \)-period problem defined in §2, the game solution starts with the two firms’ best order quantities in period \( T \), then in period \( T - 1 \), and backward period by period until period 1. It is in this regard that this game is referred to as a feedback game, and the equilibrium of the game is called the feedback Nash equilibrium (Basar and Olsder 1995):

**Definition 1.** For the two-firm game model defined by (11), the set of decisions or strategies \((Q^*_1, Q^*_2)\) constitute a (pure) feedback Nash equilibrium if and only if functions \( \pi^*_i(t) \), \( i = 1, 2; t = 1, 2, \ldots, T \), exist and satisfy the following recursive relations:

\[
\begin{align*}
\pi^*_i(t; \mu_i, \mu_{-i}) &= \max_{\mu_i} \left\{ h_i(q^*_i, \mu_i) + \gamma \pi^*_i, i+1 \right\} \\
&= h_i(q^*_i, \mu_i) + \gamma \pi^*_{i+1}(g_i(q^*_i, \mu_i, \mu_{-i})), \\
&= h_i(q^*_i, \mu_i) + \gamma \pi^*_{i+1}(g_i(q^*_i, \mu_i, \mu_{-i})), \tag{12}
\end{align*}
\]

where for \( t = T, T - 1, \ldots, 1, \)

\[
\begin{align*}
\beta^*_i &= \frac{\gamma a_i A_{i, t+1} + r_i}{\gamma a_i A_{i, t+1} + r_i + c_i}, \tag{16} \\
A_{i, t} &= \gamma A_{i, t+1} + r_i + a(\gamma a_i A_{i, t+1} + r_i + c_i)y_0(\beta^*_i) - \gamma a_{i}k_0(\beta^*_i)A_{i, t+1}, \tag{17} \\
B_{i, t} &= \gamma B_{i, t+1} + \gamma a_{i}[b_k(\beta^*_i) + k_0(\beta^*_i)]A_{i, t+1} + b(\gamma a_i A_{i, t+1} + r_i + c_i)y_0(\beta^*_i), \tag{18}
\end{align*}
\]

and for \( t = 1, 2, \ldots, T, \)

\[
\mu_{i, t+1} = [1 - \alpha_i k_0(\beta^*_i)]\mu_{i, t} + \alpha_i k_0(\beta^*_i)\mu_{i, t} - \alpha_i b_k(\beta^*_i) + \alpha_i b_k(\beta^*_i). \tag{19}
\]

### 4. Duopoly Analysis

The parameter \( A_{it} \) plays a key role in this two-firm game. Similar to the standard newsvendor problem, the optimal order quantity in period \( t \) for the dynamic competitive newsvendor problem is defined by a “critical fractile” \( \beta^*_i \). The difference here is the additional term \( \gamma a_i A_{i, t+1} \) (see Equation (16)), which measures the additional underage cost for loss of goodwill in the future. Because the market allocation \( \mu_{i, t} \) and thus the optimal order quantities are functions of \( \beta^*_i \), they are essentially determined by \( A_{it} \). Furthermore, as the coefficient of \( \mu_{i, t} \) in (15), \( A_{it} \) measures the impact of the starting market share on the corresponding firm’s total profit in the remaining planning horizon; consequently, it represents the value of market share for a competing firm. Thus, the significance of understanding \( A_{it} \) and its impact is obvious.

To analyze the behavior and the role of \( A_{it} \) for the general demand pattern, we need the following lemma:

**Lemma 2.** Let \( B_i(x) = (\gamma a_{i} x + r_i)/[(\gamma a_i x + r_i + c_i)], \tilde{S} = \{x: x \leq r_i/(ac_i)\}, \) and define, for \( x \geq 0 \)

\[
\begin{align*}
J^+_i(x) &= (1 - \gamma)x - r_i - a(\gamma a_{i} x + r_i + c_i)y_0(B_i(x)), \tag{19} \\
J^{++}_i(x) &= -r_i - a(\gamma a_{i} x + r_i + c_i)y_0(B_i(x)). \tag{20}
\end{align*}
\]

These two functions have the following properties: \( J^+_i(x) = 0 \) has a unique positive solution \( A_i^+ \); if \( S \) is not a subset of \( \tilde{S} \), \( J^{++}_i(x) = 0 \) has a unique positive solution \( A_i^{++} \). If \( S \subseteq \tilde{S} \), \( J^{++}_i(x) \) is always nonpositive and we set \( A_i^{++} = +\infty \).
Thus, the following statements are true for $A_{it}$, $A_i^*$, and $A_i^{+++}$:

1. $A_{it}^* < A_i^{+++}$.
2. If $A_{it-1}^* \in [0, A_i^*]$ for $i = 1, 2$, then $A_{it} \in [0, A_i^*]$ for $i = 1, 2$.
3. If $A_{it} \in [0, A_i^{+++})$ for $i = 1, 2$, then $A_{it+1} \in [0, A_i^{+++})$ for $i = 1, 2$.

To understand the dynamics of the two competing firms and determine whether a firm should compete aggressively, we must first determine whether the market value $A_{it}$ is positive or not. Hall and Porteus (2000) pointed out that when demands are multiplicative and $a_i < 1/2$ for all $i$, $A_{it}$ is nonnegative. With Lemma 2, we can now significantly relax the conditions for $A_{it}$ to be nonnegative.

**Theorem 2.** The value of market share $A_{it}$ is nonnegative, if either of the following two conditions is satisfied.

- **Condition 1.** $A_{it-1}^* < A_i^{+++}$, for $i = 1, 2$.
- **Condition 2.** $\alpha_i a_{it} k_i (\beta_i (A_{it}^*)) + \alpha_i a_{it} k_i (\beta_i (0)) \leq 1$, for $(i, j) = (1, 2)$ and $(i, j) = (2, 1)$.

Theorem 2 tells us that either the value of the exit market share is not too large, or the value of the current market share is increasing in the value of the next period’s market share when it is sufficiently large, the market share is always valuable in any period. It is easy to see that at least one of the two conditions is true for most practical settings. Therefore, we expect that competition will in general force every competing firm to raise its replenishment quantity to achieve a higher critical fraction than would be achieved without considering competition.

Let us consider the following condition.

- **Condition 3.** $\alpha_i a_{it} k_i (\beta_i (0)) + \alpha_j a_{it} k_j (\beta_j (0)) \leq 1$.

Condition 3 guarantees that the value of the current market share is always increasing in the value of the next period’s market share. Condition 3 implies Condition 2 in Theorem 2. By Lemma 1, Condition 3 is always satisfied if $a_i < 1/2$ for all $i$, which is the condition set in Hall and Porteus (2000). Condition 3 is very mild and can be satisfied by most parameter settings. Consider, for example, exponential (multiplicative) demands. It is easy to check that Condition 3 in this case becomes

$$\frac{\alpha_i c_i}{r_i + c_i} + \frac{\alpha_j c_j}{r_j + c_j} \leq 1.$$
From the equilibrium given in Theorem 1, we observe that \( A_{it} \) is the only time-varying parameter in the critical fractile. Thus, to determine whether a stationary equilibrium exists, we only need to check whether there is a fixed-point solution for the recursive equation that defines \( A_{it} \).

**Theorem 3.** There exists at least one fixed point \((A_{1,-\infty}, A_{2,-\infty})\) for Equation (17), with \( A_{i,-\infty} \in [0, A^*_i] \) for \( i = 1, 2 \). If the function \( l_i(x) = axk_0(\beta_i(x)) \) is nondecreasing in \( x \) for \( x \in [0, A^*_i] \), \( i = 1, 2 \), then the fixed point is unique.

For additive demands, \( l_i(x) = 0 \) and hence is non-decreasing in \( x \). For exponential demands, it is easy to check that \( l_i(x) = c_i x/ (\gamma \alpha_i x + r_i + c_i) \) is increasing in \( x \). Therefore, Theorem 3 guarantees the uniqueness of the stationary equilibrium policy when demands are either additive or exponential. We conjecture that the fixed point for Equation (17) is unique for other demand patterns, although we are currently unable to prove it without the assumption on \( l_i(x) \).

Our next question concerns the relationship between dynamic policy and stationary policy. We can demonstrate that under mild conditions, the dynamic policy will converge to the stationary policy.

**Proposition 2.** Under Condition 3, the dynamic equilibrium policy defined by Theorem 1 always converges to a stationary equilibrium policy if either of the following two conditions is satisfied:

1. \( A_{i,T+1} = 0 \), \( i = 1, 2 \).
2. The fixed point for Equation (17) is unique.

Finally, we discuss whether the competition is sustainable. We consider this question in an infinite-horizon setting to see more clearly what factors determine a firm’s survivability. For clarity, a competition is said to be sustainable for firm \( i \) if firm \( i \)’s market share is always positive throughout the game. A more general definition is to require the market share to always be no less than some predetermined positive level. This will not add theoretical difficulty and can be similarly discussed.

**Proposition 3.** For the infinite-horizon game, we define \( \Delta_i = \alpha_i k_0(\beta_i(A_{i,-\infty})) \) as firm \( i \)’s critical factor. Under Condition 3, we have:

1. The firm with a smaller critical factor can always survive; i.e., firm \( i \) can survive if \( \Delta_i \leq \Delta_j \).
2. If \( \Delta_i > \Delta_j \), firm \( i \) can survive if and only if \( b(\Delta_i/\Delta_j - 1) < a\mu_i \).
3. If the competition is sustainable for both firms, the market share for firm \( i \) in the steady state equals \( \mu_i \) if demands are additive and equals \( \mu_0 \Delta_i/(\Delta_i + \Delta_j) - b(\Delta_i - \Delta_j)/[a(\Delta_i + \Delta_j)] \) otherwise.

We observe that, essentially, different demand patterns dictate different levels of competition. Additive demands are associated with cut-throat competition, and only the strongest firms (with the smallest critical factors) will survive. That is because for additive demands, the demand variance is independent of the mean demand, and hence the coefficient of variation of the demand is decreasing in the mean demand. Therefore, the firms will aggressively compete for more market share to enjoy increasing returns to scale. In a market where brand name is highly valued, much of the potential profit is divided by a few strong firms. Without competing in price, a new firm can join the competition only by being able to attract a sufficiently large and stable customer base (through good branding). With a multiplicative market, firms can always coexist in the market and be profitable. When demands are neither additive nor multiplicative, it is possible that weak firms can survive when their critical factors are not too large. With the more general definition of survivability we have mentioned, we can conclude from Proposition 3 that the competition is sustainable for both firms if and only if their critical factors are sufficiently close to each other.

Our next question is how different parameters affect the critical factor. Consider an additive market. By Equation (21), \( A_{i,-\infty} = r_i/(1 - \gamma) \); hence the stationary critical fractile for firm \( i \) is

\[
\beta_i(A_{i,-\infty}) = r_i + \alpha_i r_i \gamma/(1 - \gamma) - r_i + \alpha_i r_i \gamma/(1 - \gamma) + c_i.
\]  

Notice that \( k_0(\beta) \) is decreasing in \( \beta \). To reduce \( \Delta_i \), we need to increase \( \beta_i \) by increasing \( r_i \) and reducing \( c_i \), as in a standard newsvendor problem. The impact of \( \alpha_i \) is more involved. We note from (22) that with larger \( \alpha_i \), firm \( i \) will raise its service level to hedge against higher switching potential, which tends to reduce \( \Delta_i \). However, the direct effect of a larger \( \alpha_i \) by the definition of \( \Delta_i \) is to increase \( \Delta_i \) proportionally. We believe that for most problem settings, the net effect of a
larger switching probability is to increase the critical factor. For multiplicative demands, we believe that the effects of the parameters are similar, but the analysis is much more complicated.

6. Summary
In this paper, we study product availability competition between two firms using a general demand model. We first show that a unique feedback Nash equilibrium strategy exists in a dynamic competition, and both the equilibrium order quantity and the corresponding discounted profit are linearly increasing in the market share of a competing firm. We also show that the value of market share depends strongly on how much the market share affects the demand variability (randomness). Firms with additive demands value their market shares more than those with multiplicative demands, because the demand variability with additive demands is independent of the market share. This qualitative understanding, presented in §4, is a major contribution of this paper.

Our second major contribution is to show that when this game is played in an infinite horizon, a stationary equilibrium always exists, and in most cases, this equilibrium is unique. We further demonstrate that under some mild conditions, the dynamic strategy always converges to a stationary policy. From both the dynamic and stationary game analyses, we can see that the demand pattern plays a major role in determining the equilibrium and the competitive behavior. Furthermore, we show that the survivability of a firm in the competition is intimately related to the demand pattern it faces. In summary, by examining this general demand pattern, we are able to develop a broader and deeper understanding of the product availability competition than currently presented in the literature.

One of the main assumptions in our model is that customers decide whether to switch to a different firm based only on their experiences with their current supplier in the previous period. This translates to a demand-switching model in which a proportion of the expected unsatisfied demand will change supplier. Any change in demand characteristics is totally determined by the change of the expectation. In practice, customers’ switching behavior in some competitive environments may be more complicated, and different demand models may be required. For example, if customers have access to all the demand information and how demands are satisfied in the previous period, we will have a very different problem and demand model. We may need a setting in which the demand variance is directly and simultaneously affected by the service level. This is obviously a challenging and interesting research topic.

Acknowledgments
The authors thank the senior editor and two anonymous referees for their detailed comments and constructive suggestions, which helped them significantly in improving both the content and the exposition of this paper. This research is supported by Hong Kong Research Council through Grant PolyU6133/02E. The authors would also like to acknowledge Zhaotong Lian for his careful reading of an earlier draft of this paper and helpful remarks.

Appendix. Proofs

Proof of Lemma 1. From Equation (1), the demand distribution function \( F(x) \) satisfies \( F(x) = F_0((x - \mu)/(\mu a + b)) \), and \( f(x) = f_0((x - \mu)/(\mu a + b)) \). Therefore, \( s_0(\beta) = \frac{s(\beta - \mu)}{a b(\mu a + b)} \) or \( s(\beta) = [a s_0(\beta) + 1] \mu + b s_0(\beta) \), and

\[
\int_0^{s(\beta)} x f(x) \, dx = \int_0^{s_0(\beta)} x f_0(x) \, dx + \int_0^{s_0(\beta)} \frac{a}{a \mu + b} \, dx + \frac{a}{a \mu + b} \int_0^{s_0(\beta)} \mu f(x) \, dx
\]

\[
=(a + b) \int_0^{s_0(\beta)} x f_0(x) \, dx + \beta \mu
\]

\[
= \left[ a \int_0^{s_0(\beta)} x f_0(x) \, dx + \beta \mu \right] + b \int_0^{s_0(\beta)} x f_0(x) \, dx.
\]

The first equation above leads to (2), while the second equation leads to (3). By the derivative rule of the inverse function, we have

\[
\frac{d}{d \beta} F_0(s_0(\beta)) = \frac{1}{f_0(s_0(\beta))}
\]

\[
y_0(\beta) = \left[ \int_{-1/a}^{s_0(\beta)} x f_0(x) \, dx \right] = s_0(\beta).
\]

From these two formulas, we can obtain \( k_0(\beta) = -(1 - \beta)/f_0(s_0(\beta)) < 0 \). Hence, \( k_0(0) = -s_0(0) < 1/a \) and \( k_0(1) = 0 \), thus \( k_0(\beta) \in [0, 1/a] \). □

Proof of Lemma 2. Function \( f^*_0(\chi) \) is strictly increasing in \( x \) for \( x \geq 0 \) because \( f^*_0(\chi) = (1 - \gamma) + \gamma_\alpha a k_0(\beta(\chi)) > 0 \). Because \( a y_0(\beta(0)) + \beta(0) \) can be treated as the coefficient of \( \mu \) in Equation (3), it must be positive; hence \( f^*_0(0) = -(r_0 + c_0) a y_0(\beta(0)) + \beta(0) < 0 \). It is also clear that \( f^*_0(\chi) = +\infty \). So \( f^*_0(\chi) = 0 \) has a unique positive solution \( \Lambda_0^* \).

Similarly, \( f^*_0(\chi) \) is also an increasing function with \( f^*_0(0) < 0 \). Applying the L’Hospital’s rule, \( f^*_0(\chi) = -r_0 +
Lemma 1 we have
\[ A_t = |1 - \gamma a_t k_0(b_t)| A_{i,t+1} - J^i_t(A_{i,t+1}). \] (23)
Here the coefficient \( \gamma a_t k_0(b_t) \in [0, 1) \), because from Lemma 1 we have \( k_0(b_t) \in [0, 1/a] \). If \( A_{i,t+1} < 0 \), then \( J^i_t(A_{i,t+1}) \leq 0 \), and we have \( A_t \geq -J^i_t(A_{i,t+1}) \geq 0 \) from Equation (23). We notice that \( A_t \) is convex in \( A_{i,t+1} \) as
\[ \frac{\partial A_t}{\partial A_{i,t+1}} = \gamma - \gamma a_t k_0(b_t) - \gamma a_t k_0(b_t) \],
\[ \frac{\partial^2 A_t}{\partial^2 A_{i,t+1}} = -\gamma^2 a_t c_i k_0(b_t) > 0. \] (24)
To show \( A_t \leq A_t^+ \), for \( A_{i,t+1} \in [0, A_t^+] \), we only need to consider the cases when \( A_{i,t+1} = 0 \) and \( A_{i,t+1} = A_t^+ \). When \( A_{i,t+1} = 0 \), \( A_t = (r_i + c_i)(a_t k_0(b_t) + b_t) \). Denote \( G_i(y) = J^i_t((r_i + c_i)(a_t k_0(b_t) + b_t)) \). It is easy to verify \( G_i(0) = 0 \) and \( G_i(y) = -A_t F_i[1 - \alpha_t k_0(b_t)(A_t)] < 0 \). Thus \( J^i_t(A_t) = G_i(y) < 0 \), which leads to \( A_t < A_t^+ \). Finally when \( A_{i,t+1} = A_t^+ \) and with \( J^i_t(A_t^+) = 0 \), it is obvious from Equation (23) that \( A_t \leq A_t^+ \). Statement (3) can be shown similarly.

Proof of Theorem 2. Using Statement (3) of Lemma 2 repeatedly, we can see that \( A_t \in [0, A_t^+] \) for all \( i \) and \( t \) if Condition 1 is true. Suppose Condition 1 does not hold, that is, \( A_{i,t+1} \geq A_t^+ \) for some finite \( A_t^+ \), but Condition 2 holds. From Equation (24) (and noting that \( k_0(b_t) \) is decreasing in \( b_t \) and hence \( k_0(b_t)(x) \) is decreasing in \( x \)), Condition 2 implies that, given fixed \( A_{i,t+1} \geq 0 \), \( A_t \) is increasing in \( A_{i,t+1} \) for \( A_{i,t+1} \geq A_t^+ \). Thus we again are sure of \( A_t \geq 0 \) for all \( i \) and \( t \).

Proof of Proposition 1. When \( t = T + 1 \), the terminal parameter \( A_{i,T+1} \) is prior and thus nonincreasing. Suppose the statement is true for \( t + 1 \). Then in period \( t \), we have
\[ \frac{dA_t}{da} = \frac{\partial A_t}{\partial A_{i,t+1}} \frac{da}{da} + (\gamma a_t A_{i,t+1} + r_i + c_i) y_t(b_t) \]
\[ - \gamma a_t k_0(b_t) A_{i,t+1}. \]
Condition 3 guarantees that \( A_t \) is nonnegative for any \( i \) and \( t \), and \( A_t \) is increasing in \( A_{i,t+1} \); hence \( \partial A_t \partial A_{i,t+1} > 0 \). We also have \( dA_t da \leq 0 \) by assumption. Thus, the first term on the right-hand side of the above equation is nonpositive. By the definition of \( y_t(b_t) \), we can see that the second term is negative. By Lemma 1, \( k_0(b_t) \geq 0 \) and \( k_0(b_t) \leq 0 \). Obviously, \( \partial b_t \partial A_{i,t+1} > 0 \), and \( dA_t da \leq 0 \) by assumption. Thus, the last term is also nonpositive. Therefore, we have shown, by induction, that \( dA_t da \leq 0 \) for all \( i \) and \( t \). Other sensitivity results can be similarly discussed.

Proof of Theorem 3. From Statement (2) of Lemma 2, Equation (17) defines a continuous mapping from a compact set \([0, A_t^+] \times [0, A_t^+] \) to itself. By Brouwer's Theorem (see, for example, Border 1985), this mapping guarantees the existence of at least one fixed point \((A_{i,\infty}, A_{j,\infty}) \in [0, A_t^+] \times [0, A_t^+] \).

From Equation (23), there is no fixed point outside the region \([0, A_t^+] \times [0, A_t^+] \). We only consider \([0, A_t^+] \times [0, A_t^+] \) for fixed points. From Equation (23), we can see that any fixed point \((A_{i,\infty}, A_{j,\infty}) \) should satisfy:
\[ -J_t^i(A_{j,\infty}) = \gamma a_t k_0(b_t)(A_{i,\infty}) A_{j,\infty} \]
\[ = \gamma a_t I_t(A_{i,\infty}) A_{j,\infty}/A_{j,\infty}. \] (25)
for \( i \neq j \). When \( a = 0 \), it is clear that the fixed point is unique and given by \((A_t^+, A_t^+)\). In the following, we consider the case with \( a > 0 \). Suppose a different fixed point exists. We can define it as \((A_{i,\infty}, A_{j,\infty}) = (\theta_1 A_{i,\infty}, \theta_2 A_{j,\infty}) \) for \( \theta_t > 1 \) without loss of generality. It is clear that \( 0 \leq -I_t^i(A_{j,\infty}) < -I_t^j(A_{i,\infty}) \), as \( I_t^i(x) \) is a strictly increasing function. For the first equality of (25) to hold for both fixed points with \((i, j) = (1, 2) \), we require \( k_0(b_t)(A_{i,\infty}) < k_0(b_t)(A_{j,\infty}) \), which is equivalent to \( \theta_1 > \theta_2 \). The second inequality comes from our assumption that \( I_t^i(x) \) is nondecreasing in \( x \) for \( x \in [0, A_t^+] \) and the fact that \((A_{i,\infty}, A_{j,\infty}) \). Therefore, we have \( \theta_1 > \theta_2 > 1 \). Similarly, Equation (25) with \((i, j) = (2, 1) \) should also hold for these two fixed points, but this will lead to \( \theta_1 > \theta_2 > 1 \) a contradiction. Therefore, the fixed point must be unique.

Proof of Proposition 2. First, we discuss the implications of Condition 3 on the mappings defined by Equation (17). Consider two mappings: \((A_{i,t+1}, A_{j,t+1}) \) to \((A_{i,t+1}, A_{j,t+1})\) and \((A_{i,t+1}, A_{j,t+1}) \) to \((A_{j,t+1}, A_{i,t+1})\). Condition 3 implies the following facts: (1) If \( 0 \leq A_{i,t+1} \leq A_t^+ \leq A_{j,t+1} \leq A_t^+ \), then \( A_t \leq A_{i,t+1} \leq A_{j,t+1} \leq A_t^+ \), and (2) if \( 0 \leq A_{i,t+1} \leq A_t^+ \leq A_{j,t+1} \leq A_{i,t+1} \leq A_{j,t+1} \), then \( A_t \leq A_{i,t+1} \) and \( A_t \leq A_{j,t+1} \). We first prove fact (1). Under Condition 3, Equation (24) implies that \( A_{i,t+1} \) is increasing in \( A_{j,t+1} \) for fixed \( A_{i,t+1} \). Thus, \( A_t \leq A_{i,t+1} \). Condition 3 also implies that \( A_t \) is nonnegative for any \( i \) and \( t \). Because \( k_0(b_t)(x) \) is decreasing in \( x \), we can see from (17) that \( A_t \) is increasing in \( A_{i,t+1} \) for fixed \( A_{j,t+1} \). This shows \( A_t \leq A_{i,t+1} \). To prove fact (2), we consider the pairs \((A_{i,t+1}, A_{j,t+1}), (A_{j,t+1}, A_{i,t+1})\) and \((A_{i,t+1}, A_{j,t+1}), (A_{j,t+1}, A_{i,t+1})\). Applying fact (1) twice, we can see fact (2) is also true.
Now, we prove the proposition under Condition (I). We use induction to show that \( A_i^t \geq A_i^{t+1} \) for all \( i \). With \( A_1, T = 0 \) and Statement (2) of Lemma 2, this is obviously true for \( t = T \). Suppose the above relation is true for some \( t \). Applying fact (2) to the pair \( (A_i, A_j) \) and \( (A_i, A_j) \), we have \( A_i, T = 1 \geq A_i^{t+1} \) for all \( i \). Therefore, the sequence \( A_i, t = 1 \), is nondecreasing and bounded from above by \( A_i^+ \) (by Lemma 2). By the monotone convergence theorem, \( A_i, T = 1 \), will converge to some constant. Clearly, the limiting vector is a fixed point of Equation (17).

Finally, we assume that Condition (2) is satisfied. Just consider nonadditive demands. Without loss of generality, we further assume that \( A_i, T = 1 \) for \( i = 1 \). Otherwise, \( A_i, T = 1 > A_i^+ \) for some \( i \); together with Equation (23), this implies that \( A_i \in [0, A_i^+] \) for some \( t \) with \( i = 1 \). We construct two sequences \( (A_i^{t+1}, A_i^{t+1}) \), and \( (A_i^{t+1}, A_i^{t+1}) \), defined by Equation (17) with \( A_i^{t+1} = 0 \) and \( A_i^{t+1} = A_i^+ \) for \( i = 1 \). We have shown that the sequence \( (A_i^{t+1}, A_i^{t+1}) \), converges to the fixed point of Equation (17). From the same discussion, we can show that the sequence \( (A_i^{t+1}, A_i^{t+1}) \), will also converge to a fixed point. From the inequalities \( A_i^{t+1} \leq A_i, T = 1 \leq A_i^{t+1} \), for \( t = 1 \), and fact (2), we have \( A_i, T = 1 \leq A_i, T = 1 \leq A_i^{t+1} \) for all \( i \) and \( t = 1 \). Take the limit \( t \rightarrow +\infty \) and notice that \( \lim_{t \rightarrow +\infty} A_i, T = 1 \) = \( \lim_{t \rightarrow +\infty} A_i, T = 1 \) = \( \lim_{t \rightarrow +\infty} A_i^{t+1} \), for \( i = 1 \), we can conclude that the sequence \( (A_i^{t+1}, A_i^{t+1}) \) will converge to the unique fixed point of Equation (17).

Proof of Proposition 3. With both firms adopting the stationary equilibrium policy, the market share is reallocated according to \( \mu_i = 1 - a\Delta_i - a\Delta_0 + b\Delta_i + b\Delta_0 \).

For additive demands, \( a = 0 \). If \( \Delta = \Delta_i \), the market allocation will be time invariant and the two firms will always keep their starting market allocation. If \( \Delta_i > \Delta_0 \), then firm \( i \) will expect to lose \( b(\Delta_i - \Delta_0) \) customers in each period until it gets ruined. These show Statements (1) to (3) for additive demands.

When demands are nonadditive, \( a > 0 \). By Condition 3, \( 1 - a\Delta - a\Delta_0 \in (0, 1) \). Therefore, \( \mu_i \) will monotonically converge to \( \mu_i = (\Delta + \Delta_0)/(\Delta + \Delta_0) - b(\Delta - \Delta_0)/[a(\Delta + \Delta_0)] \), which is positive if and only if either \( \Delta_i \leq \Delta_0 \), or otherwise \( b\Delta_i/\Delta - 1 < a\mu_i \). These confirm the proposition for nonadditive demands.

References


