2012 American Control Conference
Fairmont Queen Elizabeth, Montréal, Canada
June 27-June 29, 2012

Learning in Infinite-Horizon Inventory Competition with Total Demand Observations

Ashkan Zeinalzadeh and Aydın Alptekinoğlu and Gürdal Arslan

***Abstract*— We consider single-period and infinite-horizon inventory competition between two firms that replenish their inventories as in the well-known newsvendor model. Normally customers have a preference for shopping in one firm or the other. A fixed percentage of them who encounter a stockout in the firm of their first choice, though, visits the other firm. This substitution behavior makes the firm's replenishment decisions strategically related. Our main contribution is to introduce a simple learning algorithm to inventory competition. The learning algorithm requires each firm (a) to have the knowledge of its own critical fractile, which the firm can calculate using the values of its own per unit revenue, order cost, and holding cost; and (b) to observe its own total demand realizations. They do not necessarily know their true demand distributions. The firms need not even have any information about each other, beyond the implicit information encoded in their own total demand realizations affected by their competitors' inventory decisions. In fact, the firms need not even be aware that they are engaged in inventory competition. We prove that the inventory decisions generated by the learning algorithm converge, with probability one, to certain threshold values that constitute an equilibrium in pure Markov strategies for an infinite-horizon discounted-reward inventory competition game.**

I. INTRODUCTION

Inventory competition has received significant interest in the operations management literature. This body of work focuses on the equilibrium among firms' inventory decisions for substitutable products [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12]. For a comprehensive review, the reader is referred to [10] and [11].

In this paper we introduce a learning element to two-firm inventory competition. Consider the classic model with two substitutable products each sold by a firm. The firms make inventory replenishment decisions for their own product. Customers normally shop at either one of these firms. However, if they encounter a stockout in the firm of their first choice, a fixed percentage of them visit the other firm, which makes the firm's inventory replenishment decisions strategically related.

Our main distinction from the inventory competition literature is that we allow firms that are neither perfect decision makers nor are they fully informed about the competition they are engaged in. They make their inventory decisions using a simple learning algorithm in hopes of maximizing their expected discounted profit over the remaining periods. The learning algorithm requires each firm (a) to have the knowledge of its own critical fractile, which the firm can calculate using the values of its own per unit revenue, order cost, and holding cost; and (b) to observe its own total demand realizations. They do not necessarily know their true demand distributions. The firms need not even have any information about each other, beyond the implicit information encoded in their own total demand realizations affected by their competitors' inventory decisions. In fact, the firms need not even be aware that they are engaged in inventory competition. We prove that the inventory decisions generated by the learning algorithm converge, with probability one, to certain threshold values that constitute an equilibrium in pure Markov strategies for an infinite-horizon discounted-reward inventory competition game.

We use the following notation throughout the paper:

$:=$	defined as
$\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_{++}$	reals, nonnegative reals, positive reals
$\mathbb{R}^n$, $\mathbb{R}^n_+$, $\mathbb{R}^n_{++}$	$\underbrace{\mathbb{R}\times\cdots\times\mathbb{R}}_{n-\text{times}}$, $\underbrace{\mathbb{R}_+\times\cdots\times\mathbb{R}_+}_{n-\text{times}}$, $\underbrace{\mathbb{R}_{++}\times\cdots\times\mathbb{R}_{++}}_{n-\text{times}}$
F_X	cumulative distribution function (cdf) of X
f_X	probability density function (pdf) of X
$P[\cdot]$, $E[\cdot]$	probability, expectation
$y \leq z$ or $z \geq y$	$z-y\in\mathbb{R}^n_+$ for $y,z\in\mathbb{R}^n$
$y < z$ or $z > y$	$z-y\in\mathbb{R}^n_{++}$ for $y,z\in\mathbb{R}^n$
$\mathbf{0}$, $\mathbf{1}$	all-zeros vector in $\mathbb{R}^2$, all-ones vector in $\mathbb{R}^2$
$\Vert\cdot\Vert_p$	p-norm in $\mathbb{R}^n$, i.e., $\Vert x\Vert_p = \left(\sum_{i=1}^n \vert x_i\vert^p\right)^{1/p}$ for $x\in\mathbb{R}^n$
$(\cdot)^+$	positive part, i.e., $(x)^+ := \max\{0,x\}$ for $x\in\mathbb{R}$
$\rightarrow$	converges to
$A \Rightarrow B$	A implies B

II. MODELS

In this section we describe single- and infinite-period inventory competition games, which are both standard in the literature. We also give characterizations of equilibria in these games using existing results. What follows is a special case of the model with fixed order costs, which we describe in [12]

A. A Single-Period Inventory Competition Model

We start with two firms labeled as firm 1 and firm 2. Each firm i makes a one-time inventory level decision y_i satisfying $y_i \geq x_i$ where $x_i \geq 0$ denotes firm i's inventory level before its decision y_i is implemented. Thus, the amount of goods ordered by firm i equals $y_i - x_i$ which results in a total cost of $c_i(y_i - x_i)$ where $c_i \geq 0$ denotes the per unit cost.

This work was supported by NSF Grant #ECCS-0547692.

A. Zeinalzadeh is with Department of Electrical Engineering, University of Hawaii at Mānoa, 2540 Dole Street, Honolulu, HI 96822, USA, ashkan@hawaii.edu

A. Alptekinoğlu is with Edwin L. Cox School of Business, Southern Methodist University, 6212 Bishop Boulevard, Dallas, TX 75275, USA, aalp@smu.edu

G. Arslan is with Department of Electrical Engineering, University of Hawaii at Mānoa, 2540 Dole Street, Honolulu, HI 96822, USA, gurdal@hawaii.edu

978-1-4577-1094-0/12/$26.00 ©2012 AACC

Subsequently, each firm i faces a demand $d_i \geq 0$ caused by the buyers who prefer firm i over firm j as their first choices. If firm i cannot meet the demand d_i, in other words $d_i > y_i$, then a constant fraction $\alpha_i \in [0,1]$ of the unmet demand $d_i - y_i$ goes to the other firm j. If firm j has positive inventory after satisfying its loyal customers, i.e., $y_j > d_j$, then firm j attempts to meet the demand $\alpha_i(d_i - y_i)^+$ switching from firm i. The part of the demand $\alpha_i(d_i - y_i)^+$ that cannot be met by firm j's remaining inventory $(y_j - d_j)^+$ drops out of the system. In short, each firm i's total demand is given as [1]

$$\bar{d}_i(y_j) := d_i + \alpha_j(d_j - y_j)^+. \tag{1}$$

Once the inventory decision (y_1, y_2) are made and the total demands $(\bar{d}_1(y_2), \bar{d}_2(y_1))$ are received, each firm i generates a total revenue of $r_i s_i$ where $r_i \geq 0$ denotes the per unit price and $s_i := \min(y_i, \bar{d}_i)$ denotes firm i's sales. If a firm i has excess inventory after receiving $\bar{d}_i$, i.e., $y_i > \bar{d}_i$, then firm i pays a storage cost of $h_i(y_i - \bar{d}_i)^+$, where $h_i \geq 0$ is the per unit storage cost. In sum, firm i's net profit is given as

$$r_i \min(y_i, \bar{d}_i) - h_i(y_i - \bar{d}_i)^+ - c_i(y_i - x_i).$$

We will write each firm i's expected profit as

$$g_i(y_i, y_j; x_i) := E\left[r_i \min(y_i, \bar{d}_i) - h_i(y_i - \bar{d}_i)^+\right] - c_i(y_i - x_i).$$

Firm i's expected profit can be decomposed as

$$g_i(y_i, y_j; x_i) = \bar{g}_i(y_i, y_j) + c_i x_i \tag{2}$$

where $\bar{g}_i(y_i, y_j)$ denotes the expected profit when $x_i = 0$, and it is given as

$$\begin{aligned} \bar{g}_i(y_i, y_j) &= E\left[r_i \min(y_i, \bar{d}_i) - h_i(y_i - \bar{d}_i)^+\right] - c_i y_i \quad (3) \\ &= (r_i - c_i) y_i - (r_i + h_i) E\left[(y_i - \bar{d}_i)^+\right]. \end{aligned}$$

For fixed $x := (x_1, x_2)$, the above setup leads to a two-player single-shot simultaneous-move game with the strategy sets $\{[x_1, \infty), [x_2, \infty)\}$ and utility functions given in (2). This game will be referred to as Γ^x. The prevailing notion of solution in game theory has been that of (Nash) equilibrium. A pair of inventory decisions (y_1^*, y_2^*) for firm 1 and firm 2, respectively, constitute an equilibrium of the game Γ^x if and only if

$$y_i^* \geq x_i \quad \text{and} \quad \bar{g}_i\left(y_i^*, y_j^*\right) = \max_{y_i \geq x_i} \bar{g}_i\left(y_i, y_j^*\right), \qquad \text{for all } i.$$

An equilibrium strategy (y_1^*, y_2^*) is known as a person-by-person optimal solution, that is, y_i^* is optimal for firm i as long as the other firm j plays its equilibrium strategy y_j^*.

The single-period inventory competition model introduced above is consistent with the existing literature on this subject.

B. An Infinite-Horizon Inventory Competition Model

Here, we assume that the single-period inventory competition game described in the previous subsection is repeatedly played by two firms infinitely many times. Each firm i starts with some given inventory level $x_{i,1} \geq 0$. Then, the following events take place within any period $t \in \{1, 2, \dots\}$. First, each firm i makes an inventory decision $y_{i,t}$ satisfying $y_{i,t} \geq x_{i,t}$, where $x_{i,t}$ denotes firm i's inventory carried from the previous period. Second, the demand $d_{i,t}$ from firm i's loyal customers is realized. Third, a fixed fraction $\alpha_i \in [0,1]$ of the unmet demand at each firm i goes to the other firm j. This results in a total demand of $\bar{d}_{i,t}(y_{j,t}) = d_{i,t} + \alpha_j(d_{j,t} - y_{j,t})^+$ for each firm i. Next, each firm i makes some profit (or loss) with the expected profit being equal to $g_i(y_{i,t}, y_{j,t}; x_{i,t})$, given $x_{i,t}$. Last, the excess inventory $x_{i,t+1} = \left(y_{i,t} - \bar{d}_{i,t}(y_{j,t})\right)^+$ of each firm i is stored and carried to the next period.

The repeated inventory competition described above leads to an infinite-horizon dynamic game. We only consider pure Markov strategies where each firm i's decision $y_{i,t}$ in period t depends on $(x_{i,t}, x_{j,t})$ (the dependence of $y_{i,t}$ on $x_{j,t}$ will be removed later). More precisely, each firm i's decision $y_{i,t}$ in period t is given by $y_{i,t} = \eta_{i,t}(x_{i,t}, x_{j,t})$ where $\eta_i := (\eta_{i,1}, \eta_{i,2}, \dots)$ is firm i's (pure Markov) strategy which satisfies

$$\eta_{i,t}(x_{1,t}, x_{2,t}) \geq x_{i,t}, \qquad \text{for all } t \in \{1, 2, \dots\}. \tag{4}$$

If each firm i uses such a strategy η_i, then firm i's expected discounted profit, with the discount factor $\delta_i \in (0,1)$, is given as

$$J_i(\eta_i, \eta_j) = \lim_T E\left[\sum_{t=1}^{T} \delta_i^{t-1} g_i(y_{i,t}, y_{j,t}; x_{i,t})\right] \tag{5}$$

where $y_{i,t} = \eta_{i,t}(x_{i,t}, x_{j,t})$, $x_{i,t+1} = \left(y_{i,t} - \bar{d}_{i,t}\right)^+$, and the expectation above is taken over $\{(x_{1,t}, x_{2,t})\}_{t \in [1,T]}$. We will call a pair of strategies (η_1^*, η_2^*) satisfying (4) an equilibrium if and only if, for all i,

$$J_i(\eta_i^*, \eta_j^*) = \max\left\{J_i(\eta_i, \eta_j^*) : \eta_i \text{ satisfies (4)}\right\}.$$

We should point out that, in order to make a decision in period t, a firm i using a pure Markov strategy η_i of the form introduced above needs access to not only $x_{i,t}$ but also $x_{j,t}$. This is clearly an excessive requirement on each firm. In the next section, we will present an equilibrium strategy which does not require any firm i to have access to $x_{j,t}$.

The following conditions are assumed to hold in the remainder of the paper for simplicity and to avoid trivialities.

Assumption 1: For all $i \in \{1, 2\}$,

1) $x_i \geq 0$, $x_{i,1} \geq 0$, $c_i \geq 0$, $h_i \geq 0$, $\delta_i \in (0,1)$, $\alpha_i \in [0,1)$
2) $r_i > c_i$
3) $c_i + h_i > 0$
4) $d = (d_1, d_2)$ is a random vector in $\mathbb{R}_+^2$ with a continuous pdf f_d satisfying: for all $y \in \mathbb{R}_{++}^2$, $f_d(y) > 0$.
5) $\{(d_{1,t}, d_{2,t})\}_{t \geq 1}$ is iid with the common cdf F_d.

Part (1) of Assumption 1 rules out unreasonable cases ($\alpha_i < 1$ is not necessary but it is assumed so that we can use some existing results in the literature). Part (2) is made to avoid the trivial case where both firms do not order any goods under any circumstance. Part (3) is made to avoid another trivial case where the goods are ordered and stored for free, therefore, each firm can order any amount of goods with no cost in which case there would never be any unmet demand.

[1] The dependence of $\bar{d}_i$ on y_j will not always be made explicit.

Part (4) is needed for the contraction property of the players' best response functions (Appendix A). The contraction property leads to the existence of a unique equilibrium that can be obtained via the best response dynamics. Part (4) can be relaxed to accommodate, in particular, some demand distributions with bounded support. Part (5) is used to show the convergence of our learning algorithm.

III. Equilibrium Strategies in Inventory Competition

In this section, we review the equilibrium strategies, in particular, the issue of existence of equilibrium strategies in inventory competition models. We cite some results from the literature for both the single-period and the infinite-horizon inventory competition models.

A. The Single-Period Case

We first define the best response function for firms taking part in the game Γ^x as follows: $BR^x = (BR^x_1, BR^x_2)$ where, for all i and $y_j \in \mathbb{R}_+$,

$$BR^x_i(y_j) \quad := \quad \mathrm{argmax}_{y_i \geq x_i} \bar{g}_i(y_i, y_j). \tag{6}$$

For notational convenience, we use BR to denote BR^x when $x = \mathbf{0}$. Clearly, a profile $y^* \in \mathbb{R}^2_+$ of inventory replenishment decisions is an equilibrium of Γ^x if and only if $y^* = BR^x(y^*)$. We call $y^* \in \mathbb{R}^2_+$ an *interior* equilibrium of Γ^x when all firms place orders in equilibrium, i.e., $x < y^* = BR^x(y^*)$.

The existence and uniqueness of equilibrium have been studied in the literature for the case of zero initial inventory levels, i.e., $x = \mathbf{0}$. It has been shown that $\Gamma^{\mathbf{0}}$ always possess an equilibrium; moreover, the uniqueness of equilibrium has been shown under mild conditions; see [1] for two firms, [13] for three firms, [14], [8] for arbitrary number of firms, [2], [5], [10] for generalizations of demand generation and substitution models. The following extension to the case of $x \neq \mathbf{0}$ readily follows from Proposition 4 in [8].

Proposition 1: Let Assumption 1 hold.

1) Γ^x possesses a unique equilibrium which can be obtained by the best response iterations $y^{\ell+1} = BR^x\left(y^\ell\right)$.
2) If y^* is an equilibrium of $\Gamma^{\mathbf{0}}$ and $x \leq y^*$, then y^* is an equilibrium of Γ^x.
3) If y^* is an equilibrium of Γ^x and $x < y^*$, then y^* is an equilibrium of $\Gamma^{\mathbf{0}}$.

Note that, for an interior equilibrium with both firms placing orders to occur, the initial inventories must be below an equilibrium of $\Gamma^{\mathbf{0}}$.

B. The Infinite-Horizon Case

Here, we review some results from [10]. In [10], it is shown that equilibrium strategies, which are in the form of constant threshold strategies, can be constructed for firms engaged in any infinite-horizon discounted-reward inventory competition. The approach in [10] is to rearrange the terms in (5) and rewrite it as

$$\begin{aligned}\lim_T E&\left[\sum_{t=1}^{T} \delta_i^{t-1} g_i\left(y_{i,t}, y_{j,t}; x_{i,t}\right)\right] \\ &= \lim_T E\left[c_i x_{i,1} + \sum_{t=1}^{T} \delta_i^{t-1} \tilde{g}_i(y_{i,t}, y_{j,t})\right]\end{aligned} \tag{7}$$

where, for all i and $y \in \mathbb{R}^2_+$,

$$\tilde{g}_i(y) := (r_i - c_i)\, y_i - (r_i + h_i - \delta_i c_i)\, E\left[\left(y_i - \bar{d}_i\left(y_j\right)\right)^+\right]. \tag{8}$$

Now, consider a single-period inventory competition model $\tilde{\Gamma}^x$ where firm i's expected profit function is given by (8), i.e., $\tilde{\Gamma}^x$ is obtained from Γ^x by replacing h_i with $h_i - \delta_i c_i$. Let $\tilde{y}$ be the equilibrium of $\tilde{\Gamma}^{\mathbf{0}}$ (which exists by Proposition 1), and let the strategy $\tilde{\eta}$ be defined by: for all i and t,

$$\tilde{\eta}_{i,t}\left(x_{i,t}, x_{j,t}\right) := \max\left\{\tilde{y}_i, x_{i,t}\right\}. \tag{9}$$

The strategy $\tilde{\eta}$ is a threshold strategy; it satisfies (4), and requires firm i to observe $x_{i,t}$, but not $x_{j,t}$, in period t. A little thought reveals that, if $(x_{1,1}, x_{2,1}) \leq \tilde{y}$, then $\tilde{\eta}$ is also an equilibrium strategy for the infinite-horizon inventory competition where firm i's expected discounted profit is (7). This is because 1) there is no inter-period dependency in (7), other than the constraint $y_{i,t} \geq x_{i,t} = \left(y_{i,t-1} - \bar{d}_{i,t-1}\left(y_{j,t-1}\right)\right)^+$, and 2) $\tilde{y}$ "equilibrates" (7) period by period while satisfying the constraint $\tilde{y}_i \geq x_{i,t}$ when $\tilde{y}_i \geq x_{i,1}$.

Proposition 2: Let Assumption 1 hold. If $(x_{1,1}, x_{2,1}) \leq \tilde{y}$, then $\tilde{\eta}$ defined by (9) is an equilibrium strategy for the infinite-horizon inventory competition model.

Having discussed the basics of equilibria, we move on to the issue of how an equilibrium can arise in an actual play of an inventory competition game involving imperfect firms.

IV. Learning in Infinite-Horizon Inventory Competition with Total Demand Observations

Traditionally, an equilibrium is justified as the predicted outcome of a noncooperative game if it is "common knowledge" that all players are rational and know the utility functions of their own as well as their opponents'; see [15], [16]. In contrast, an equilibrium can also be justified if it emerges as the long-term outcome of an iterative procedure whereby players with limited rationality and information grope for individual optimality by making repeated decisions based on their observations of the past play. This is the essence of what is known as the "learning in games" approach for which we refer the reader to the books [17], [18], [19], [20], [21] and the references therein.

Similarly, this paper takes a learning approach to infinite-horizon inventory competition and deals with the question of whether or not firms can learn to play an equilibrium strategy where each firm 1) observes only its own past decisions and total demands, and 2) knows only its own critical ratio $\tilde{\gamma}_i :=$

$\frac{r_i - c_i}{r_i + h_i + c_i}$, in particular without any knowledge of any firm's demand distributions including its own.

Accordingly, we consider a learning scenario in which each firm i chooses its inventory level in each period to approximately maximize the expected profit function (8) based on its observation history. First, suppose that firm i wishes to maximize $\tilde{g}_i(\cdot, y_{j,t+1})$ in period $t+1$ and somehow it has access to $F_{\bar{d}_i(y_{j,t+1})}$ before making its decision $y_{i,t+1}$ in period $t+1$. Then, in view of Appendix A, the optimal choice for firm i would be

$$y_{i,t+1} = \max\left\{x_{i,t+1}, F^{-1}_{\bar{d}_i(y_{j,t+1})}(\tilde{\gamma}_i)\right\}. \qquad (10)$$

In actuality, firm i does not have access to $F_{\bar{d}_i(y_{j,t+1})}$ at all. However, firm i has access to its history of total demand realizations $\left(\bar{d}_{i,1}(y_{j,1}), \ldots, \bar{d}_{i,t}(y_{j,t})\right)$, in period $t+1$. Therefore, we assume that firm i replaces $F_{\bar{d}_i(y_{j,t+1})}$ in (10) with the sample distribution defined by

$$F_{i,t}(\xi) := \frac{1}{t}\sum_{k=1}^{t} I\left\{\bar{d}_i(y_{j,k}) \leq \xi\right\}, \quad \text{for } \xi \in \mathbb{R}.$$

More precisely, each firm i updates its inventory level according to

$$y_{i,t+1} = \min\{\xi \geq x_{i,t+1} : F_{i,t}(\xi) \geq \tilde{\gamma}_i\}, \qquad y_{i,1} \geq x_{i,1}. \quad (11)$$

In the literature, the right hand side of the equality in (11) is called t-th sample quantile at $\tilde{\gamma}_i$, when $x_{i,t+1} = 0$. Note that, if $\bar{d}_{i,k:t}$ denotes the $k-$th smallest value of $\bar{d}_{i,1}, \ldots, \bar{d}_{i,t}$, then we can write (11) as

$$y_{i,t+1} = \max\left\{x_{i,t+1}, \bar{d}_{i,\lceil\tilde{\gamma}_i t\rceil:t}\right\}, \qquad y_{i,1} \geq x_{i,1} \qquad (12)$$

where $\lceil\cdot\rceil$ denotes the integer ceiling. If firm i generates its inventory levels $y_{i,t}$ by (12) and firm i's competitor is constant at some $\hat{y}_j \in \mathbb{R}_+$, i.e., $y_{j,t} = \hat{y}_j$, for all $t \geq 1$, then the convergence of $y_{i,t}$ to an optimal inventory level with probability one can be obtained by the existing results in the literature on the convergence of sample quantile processes, because $\bar{d}_{i,1}, \bar{d}_{i,2}, \ldots$ is an iid sequence in this case; see (1.4.9) in [22]. Whereas, if both firms generate their inventory levels by (12), then $\bar{d}_{i,1}, \bar{d}_{i,2}, \ldots$ is no longer an iid sequence for any firm i. Hence, in this case, the convergence of the inventory levels does not readily follow from the existing results in the literature. Presenting such a convergence result is the main objective of this section.

We now state the main result of this paper whose proof is provided in Appendix B.

Theorem 1: Let Assumption 1 hold. The inventory decisions generated by (12) converge, with probability one, to the unique equilibrium of the single-period inventory competition model $\tilde{\Gamma}^{\mathbf{0}}$.

Recall, from Proposition 2, that a threshold strategy obtained from an equilibrium of $\tilde{\Gamma}^{\mathbf{0}}$ is an equilibrium strategy for the infinite-horizon inventory competition (if $(x_{1,1}, x_{2,1}) \leq \tilde{y}$). In view of this, Theorem 1 implies that firms generating their inventory decisions by (12) will asymptotically learn to play an equilibrium strategy in the infinite-horizon inventory competition.

A. An Illustrative Example

We now verify our main convergent result by numerical simulation of the process (12) for the following parameter values: for all i,

$$r_i = 4,\ c_i = 4,\ h_i = 1,\ \delta_i = 0.9,\ \alpha_i = 0.7.$$

Moreover, for each i, $\{d_{i,t}\}_{t\geq 1}$ is independently sampled from a random variable uniformly distributed over $[0,1]$. The equilibrium of the single-period game $\tilde{\Gamma}^{\mathbf{0}}$ in this scenario is computed as

$$(\tilde{y}_1, \tilde{y}_2) = (0.6644, 0.6644).$$

Figure 1 shows a sample path of $(y_{1,t}, y_{2,t})$ generated by the process (12) with randomly chosen initial conditions $(x_{1,1}, x_{2,1})$ and $(y_{1,1}, y_{2,1})$ for 200 steps (the constant lines show the respective equilibrium inventory levels $(\tilde{y}_1, \tilde{y}_2)$). Clearly, the sample path shown in Figure 1 indicates convergence to $(\tilde{y}_1, \tilde{y}_2)$.

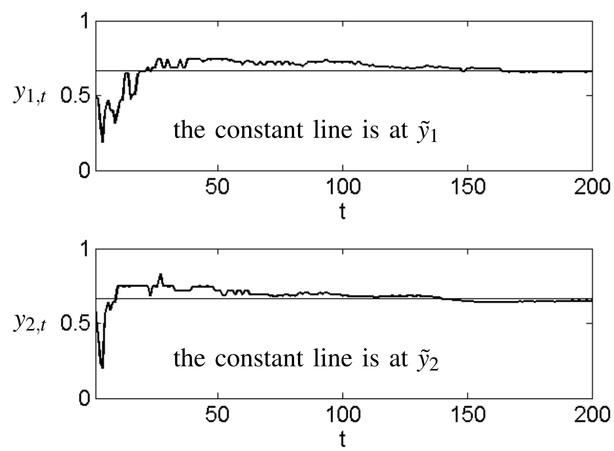

Fig. 1. A sample path of $(y_{1,t}, y_{2,t})$ generated by the process (12).

V. CONCLUSIONS

In this paper we inquire whether or not two firms can learn to play "optimally" in an infinite-horizon discounted-reward inventory competition game. We introduce a learning algorithm by which firms can make their inventory replenishment decisions in each period of an infinite-horizon inventory competition game. The learning algorithm requires each firm only to know its own critical ratio and to make observations of its own total demand realizations. It allows the firms to asymptotically play person-by-person optimal strategies, i.e., equilibrium strategies for the infinite-horizon discounted-reward inventory competition game. Our main contribution is to present a proof of convergence of the firms' decisions to such an equilibrium. Extending this work to the case with firms observing only their own sales remains as a significant research problem.

APPENDIX A: BEST RESPONSE FUNCTION

If Assumption 1 hold, BR^x satisfies the following in $\mathbb{R}_+^2$.

1) For all i and $y_j \in \mathbb{R}_+$,

$$BR_i^x(y_j) = \max\left\{x_i, F^{-1}_{\bar{d}_i(y_j)}(\gamma_i)\right\} \text{ where } \gamma_i = \frac{r_i - c_i}{r_i + h_i}. \tag{13}$$

2) BR^x is nonincreasing, i.e., $\check{y} \geq \hat{y} \Rightarrow BR^x(\check{y}) \leq BR^x(\hat{y})$.
3) BR^x is a contraction.

APPENDIX B: PROOF OF THEOREM 1

Let the firms' inventory levels $\{y_t\}_{t\geq 1}$ be generated by (12), i.e., for all i, t,

$$\begin{aligned} x_{i,t+1} &:= \left(y_{i,t} - \bar{d}_{i,t}(y_{j,t})\right)^+ \\ y_{i,t+1} &:= \min\left\{\xi \geq x_{i,t+1} : \frac{1}{t}\sum_{k=1}^{t} I\left\{\bar{d}_{i,k}\left(y_{j,k}\right) \leq \xi\right\} \geq \tilde{\gamma}_i\right\} \end{aligned} \tag{14}$$

with $y_{i,1} \geq x_{i,1}$, for given $x_{i,1} \geq 0$. Let the sequence $\left\{y_t^1\right\}_{t\geq 1}$ be identically equal to $\mathbf{0}$, and let the family of sequences $\left\{y_t^\ell\right\}_{t\geq 1}$, for $\ell = 2,3,\ldots$, be recursively generated by, for all i, t,

$$\begin{aligned} x_{i,t+1}^\ell &:= \left(y_{i,t}^\ell - \bar{d}_{i,t}\left(y_{j,t}^\ell\right)\right)^+ \\ y_{i,t+1}^\ell &:= \min\left\{\xi \geq x_{i,t+1}^\ell : \frac{1}{t}\sum_{k=1}^{t} I\left\{\bar{d}_{i,k}\left(y_{j,k}^{\ell-1}\right) \leq \xi\right\} \geq \tilde{\gamma}_i\right\} \end{aligned} \tag{15}$$

with $y_{i,1}^\ell = y_{i,1}$. For the same sequence of demand realizations $\{d_t\}_{t\geq 1}$, it is easy to see that, for all t,

$$y_t^1 \leq y_t^3 \leq \cdots \leq y_t \leq \ldots y_t^4 \leq y_t^2. \tag{16}$$

The desired result now follows from (16) and Claim 1 below.

Claim 1:

$$P\left[\lim_{\ell,t} y_t^\ell = \tilde{y}\right] = 1 \tag{17}$$

where $\tilde{y}$ is the unique equilibrium of the single-period inventory competition game $\tilde{\Gamma}^{\mathbf{0}}$.

Proof: Let $\left\{y^\ell\right\}_{\ell\geq 1}$ be generated by $y^{\ell+1} = \tilde{BR}\left(y^\ell\right)$ with $y^1 = \mathbf{0}$, where $\tilde{BR}$ is the best response function for $\tilde{\Gamma}^{\mathbf{0}}$, which can be obtained from BR, see Appendix A, by replacing γ_i with $\tilde{\gamma}_i$, for all i. From Proposition 1, we have $\lim_\ell y^\ell = \tilde{y}$. Hence, (17) will follow, if we show that, for all $\ell \geq 1$,

$$P\left[\lim_t y_t^\ell = y^\ell\right] = 1. \tag{18}$$

We show (18) by induction. Clearly, $P\left[\lim_t y_t^1 = y^1\right] = 1$. Hence, let us assume that, for some $\ell \geq 1$,

$$P\left[\lim_t y_t^\ell = y^\ell\right] = 1. \tag{19}$$

Since F_d is continuous, a generalization of the well-known Glivenko-Cantelli theorem (see Theorem 7.1 in [23]) implies, for all i,

$$P\left[\sup_{y_j\in\mathbb{R}_+}\sup_{\xi\in\mathbb{R}}\left|\frac{1}{t}\sum_{k=1}^{t} I\left\{\bar{d}_{i,k}(y_j) \leq \xi\right\} - F_{\bar{d}_i(y_j)}(\xi)\right| \to 0\right] = 1. \tag{20}$$

Also, since $E[d_i] > 0$ (due to Assumption 1), the strong law of large numbers implies, for all i,

$$P\left[\sum_{t\geq 1} d_{i,t} = \infty\right] = 1. \tag{21}$$

Now, let Ω be the set of demand realization sequences such that, for each $\omega = \{d_t\}_{t\geq 1} \in \Omega$, and for all i,

- given $\varepsilon > 0$, there exists a positive integer $t_\varepsilon(\omega)$ such that $\sup_{t\geq t_\varepsilon(\omega)}\left\|y_t^\ell - y^\ell\right\|_1 < \varepsilon$, and
- $$\sup_{y_j\in\mathbb{R}_+}\sup_{\xi\in\mathbb{R}}\left|\frac{1}{t}\sum_{k=1}^{t} I\left\{\bar{d}_{i,k}(y_j) \leq \xi\right\} - F_{\bar{d}_i(y_j)}(\xi)\right| \to 0,$$ and
- $\sum_{t\geq 1} d_{i,t} = \infty$.

In view of (19)-(21), $P[\Omega] = 1$. We have, for all $\omega \in \Omega$, $\varepsilon > 0$, $t \geq t_\varepsilon(\omega)$, i, $\xi \in \mathbb{R}$,

$$\begin{aligned} &\frac{1}{t}\sum_{k=t_\varepsilon(\omega)}^{t} I\left\{\bar{d}_{i,k}\left(y_j^\ell - \varepsilon\mathbf{1}\right) \leq \xi\right\} \\ &\quad\leq \frac{1}{t}\sum_{k=t_\varepsilon(\omega)}^{t} I\left\{\bar{d}_{i,k}\left(y_{j,k}^\ell\right) \leq \xi\right\} \\ &\quad\leq \frac{1}{t}\sum_{k=t_\varepsilon(\omega)}^{t} I\left\{\bar{d}_{i,k}\left(y_j^\ell + \varepsilon\mathbf{1}\right) \leq \xi\right\} \end{aligned}$$

which implies that

$$\begin{aligned} F_{\bar{d}_i\left(y_j^\ell - \varepsilon\mathbf{1}\right)}(\xi) &\leq \liminf_t \frac{1}{t}\sum_{k=1}^{t} I\left\{\bar{d}_{i,k}\left(y_{j,k}^\ell\right) \leq \xi\right\} \\ &\leq \limsup_t \frac{1}{t}\sum_{k=1}^{t} I\left\{\bar{d}_{i,k}\left(y_{j,k}^\ell\right) \leq \xi\right\} \\ &\leq F_{\bar{d}_i\left(y_j^\ell + \varepsilon\mathbf{1}\right)}(\xi). \end{aligned} \tag{22}$$

In addition, we always have, for all $\varepsilon > 0$, i, $\xi \in \mathbb{R}$,

$$\begin{aligned} F_{\bar{d}_i\left(y_j^\ell\right)}(\xi - \varepsilon\bar{\alpha}) &\leq F_{\bar{d}_i\left(y_j^\ell - \varepsilon\mathbf{1}\right)}(\xi) \\ &\leq F_{\bar{d}_i\left(y_j^\ell + \varepsilon\mathbf{1}\right)}(\xi) \\ &\leq F_{\bar{d}_i\left(y_j^\ell\right)}(\xi + \varepsilon\bar{\alpha}) \end{aligned} \tag{23}$$

where $\bar{\alpha} := \max_i \alpha_i$. Also, note that $F_{\bar{d}_i\left(y_j^\ell\right)}$ is uniformly continuous[2]. Therefore, from (22)-(23), we have, for all $\omega \in \Omega$, i,

$$\lim_t \sup_{\xi\in\mathbb{R}}\left|\frac{1}{t}\sum_{k=1}^{t} I\left\{\bar{d}_{i,k}\left(y_{j,k}^\ell\right) \leq \xi\right\} - F_{\bar{d}_i\left(y_j^\ell\right)}(\xi)\right| = 0.$$

[2]A continuous cdf is always uniformly continuous.

This, by the strict monotonicity of $F_{\bar{d}_i\left(y_j^\ell\right)}$, implies that, for all $\omega \in \Omega$, i,

$$\lim_t \min\left\{\xi : \frac{1}{t}\sum_{k=1}^{t} I\left\{\bar{d}_{i,k}\left(y_{j,k}^\ell\right) \le \xi\right\} \ge \gamma_i\right\}$$
$$= \min\left\{\xi : F_{\bar{d}_i\left(y_j^\ell\right)}(\xi) \ge \gamma_i\right\} = y_i^{\ell+1}. \qquad (24)$$

In view of (15) and (24), for each $\omega \in \Omega$, the following statement holds true: given $\delta > 0$, there exists a $\hat{t}_\delta(\omega)$ such that for all $t \ge \hat{t}_\delta(\omega)$, i,

$$y_i^{\ell+1} - \delta \le y_{i,t+1}^{\ell+1} \le \max\left\{y_i^{\ell+1} + \delta, x_{i,t+1}^{\ell+1}\right\}. \qquad (25)$$

In reference to (25), suppose that

$$y_{i,t+1}^{\ell+1} > y_i^{\ell+1} + \delta, \text{ for all } t \ge \hat{t}_\delta(\omega).$$

Then,

$$\begin{aligned} & y_{i,t+1}^{\ell+1} \le x_{i,t+1}^{\ell+1}, \text{ for all } t \ge \hat{t}_\delta(\omega) \\ \Rightarrow \quad & x_{i,t+2}^{\ell+1} \le \left(x_{i,t+1}^{\ell+1} - d_{i,t+1}\right)^+, \text{ for all } t \ge \hat{t}_\delta(\omega) \\ \Rightarrow \quad & \lim_t x_{i,t}^{\ell+1} = 0 \quad \left(\text{since } \sum_{t\ge 1} d_{i,t} = \infty\right) \end{aligned}$$

which leads to the contradiction $\lim_t y_{i,t}^{\ell+1} = 0$. Therefore, for some $\tilde{t}_\delta(\omega) \ge \hat{t}_\delta(\omega)$ and all i, it must be true that $y_{i,\tilde{t}_\delta(\omega)+1}^{\ell+1} \le y_i^{\ell+1} + \delta$, which implies $x_{i,\tilde{t}_\delta(\omega)+2}^{\ell+1} \le y_i^{\ell+1} + \delta$, which in turn implies $y_{i,\tilde{t}_\delta(\omega)+2}^{\ell+1} \le y_i^{\ell+1} + \delta$. As a result, we must have $y_{i,t+1}^{\ell+1} \le y_i^{\ell+1} + \delta$, for all $t \ge \tilde{t}_\delta(\omega)$, i. In conclusion, we must have, for all $\omega \in \Omega$, $\delta > 0$, $t \ge \tilde{t}_\delta(\omega)$, i,

$$\left|y_{i,t+1}^{\ell+1} - y_i^{\ell+1}\right| \le \delta.$$

This completes the induction and proves (18). □

REFERENCES

[1] M. Parlar, "Game theoretic analysis of the substitutable product inventory problem with random demands," *Naval Research Logistics*, vol. 35, no. 3, pp. 397–409, 1988.

[2] S. A. Lippman and K. F. McCardle, "The competitive newsboy," *Operations Research*, vol. 45, no. 1, pp. 54–65, 1997.

[3] R. Ernst and P. Kouvelis, "The effects of selling packaged goods on inventory decisions," *Management Science*, vol. 45, no. 8, pp. 1142–1155, 1999.

[4] S. Mahajan and G. van Ryzin, "Stocking retail assortments under dynamic consumer substitution," *Operations Research*, vol. 49, no. 3, pp. 334–351, 2001.

[5] ——, "Inventory competition under dynamic consumer choice," *Operations Research*, vol. 49, no. 5, pp. 646–657, 2001.

[6] K. Rajaram and C. S. Tang, "The impact of product substitution on retail merchandising," *European Journal of Operational Research*, vol. 135, no. 3, pp. 582–601, 2001.

[7] Z. M. Avsar and M. Baykal-Gursoy, "Inventory control under substitutable demand: A stochastic game application," *Naval Research Logistics*, vol. 49, no. 4, pp. 359–375, 2002.

[8] S. Netessine and N. Rudi, "Centralized and competitive inventory models with demand substitution," *Operations Research*, vol. 51, no. 2, pp. 329–335, 2003.

[9] F. Bernstein and A. Federgruen, "Dynamic inventory and pricing models for competing retailers," *Naval Research Logistics*, vol. 51, no. 2, pp. 258–274, 2004.

[10] S. Netessine, N. Rudi, and Y. Wang, "Inventory competition and incentives to back-order," *IIE Transactions*, vol. 38, no. 11, pp. 883–902, 2006.

[11] T. L. Olsen and R. P. Parker, "Inventory management under market size dynamics," *Management Science*, vol. 54, no. 10, pp. 1805–1821, 2008.

[12] A. Zeinalzadeh, A. Alptekinoğlu, and G. Arslan, "On Existence of Equilibrium in Inventory Competition with Fixed Order Costs," Dec. 2011, sumitted to *Naval Research Logistics*.

[13] M. Parlar and Q. Wang, "A three-person game theory model arising in stochastic inventory control theory," *European Journal of Operational Research*, vol. 76, no. 1, pp. 83–97, 1994.

[14] R. Karjalainen, "The newsboy game," Wharton School, University of Pennsylvania, Tech. Rep., 1992.

[15] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. Philadelphia, PA: SIAM, 1999.

[16] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.

[17] H. P. Young, *Strategic Learning and Its Limits*. Oxford University Press Inc.: SIAM, 2004.

[18] ——, *Individual Strategy and Social Structure*. Princeton, NJ: Princeton University Press, 1998.

[19] D. Fudenberg and D. Levine, *The Theory of Learning in Games*. Cambridge, MA: MIT Press, 1998.

[20] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*. Cambridge, UK: Cambridge University Press, 1998.

[21] J. Weibull, *Evolutionary Game Theory*. Cambridge, MA: MIT Press, 1995.

[22] M. Csörgö, *Quantile processes with statistical applications*, ser. Volume 42 of CBMS-NSF regional conference series in applied mathematics. SIAM, 1983.

[23] R. R. Rao, "Relations between weak and uniform convergence of measures with applications," *The Annals of Mathematical Statistics*, vol. 33, no. 2, pp. 659–680, June 1962.