

A Dynamic Queuing Model

Susheng WANG and Lijing ZHU⁺

Hong Kong University of Science and Technology

Received May 2004

This paper proposes a dynamic queuing model in which excess demand is cleared through multiple shifts. Due to differences in valuation and costs, some consumers choose to compete in queues for early consumption, while others avoid queues by late consumption. A unique efficient rational expectations equilibrium is shown to exist and some general characteristics of the queuing equilibrium are analyzed. The theory is then applied to some interesting realistic situations such as shopping, highways, and restaurants.

Key Words: queuing, capacity constraint, multiple shifts

JEL Classification: 91B26, 91B74,

⁺ Please address correspondence to: S. Wang, Department of Economics, HKUST, Clear Water Bay, Hong Kong. E-mail: *s.wang@ust.hk*.

1. Introduction

In an efficient market, the price mechanism is supposed to function perfectly to clear the market at all times. However, there are many realistic situations in which this does not happen; the price either fails to adjust to demand or simply a market for the product is missing. Instead, nonprice mechanisms such as queuing, lotteries and coupons are used as effective mechanisms to allocate resources. This paper studies queuing in a dynamic setting.

Consider a supplier who provides a product with a capacity constraint. When there is excess demand and when the price mechanism fails to react, consumers may have to compete in a queue for the allocation of the product. In existing queuing models,¹ queuing is a one-shot game in which consumers compete for a product by deciding to stand in a queue or not. In equilibrium, those who are in the queue will get the product while the rest will not. In a dynamic setting with multiple shifts, a consumer not only decides whether or not to compete in a queue, but also which shift he will compete in. For example, if a consumer is patient enough, he can avoid queuing by late consumption. Multiple shifts allow a given capacity to be used repeatedly (saving costs) and alleviate the demand for early consumption (improving welfare). We find that allowing more shifts can dramatically improve efficiency; in fact, if many consumers are patient, the efficiency of a multi-shift queuing mechanism can be fairly close to that of the price mechanism.

In our dynamic queuing model, consumers form expectations about the length of the queuing time in each shift and decide whether or not to consume at all, and if so, in which shift to consume. Under fairly general conditions, a stable and efficient rational expectations queuing equilibrium exists in which the consumers satisfy both the incentive compatibility and individual rationality constraints. With the queuing times for multiple shifts determined by valuation and costs, the consumers' effective demand, after incorporating costs of queuing and delay, is then derived. In equilibrium, the queuing time in each shift equals the queuing time determined by the effective demand and the capacity.

¹See, for example, early works by Leeman (1964), Naor (1969) and Barzel (1974), a game theoretical approach by Holt-Sherman (1982), and a model allowing differential time costs and personal valuations among consumers by Suen (1989).

This theory is readily applicable to many realistic situations in which consumers have a choice between early and late consumption. Early consumption (or, more precisely, timing consumption) gives a consumer higher satisfaction, but late consumption means less queuing. In particular, we look at three applications: shopping, highways and restaurants. There are situations in which shoppers have to queue to get certain products. Some shoppers with certain preferences may avoid queues by buying a similar or the same product for a higher price at another store. On highways, there are rush hours during which drivers have to queue for the use of highways. Some drivers with certain preferences may avoid the rush hours by using the highways at different times. Restaurants can be very crowded at certain points during a day. Some people with certain preferences may avoid those time points by adjusting their dining times slightly.

The paper proceeds as follows. Section 2 develops a theory for a dynamic queuing model. Three examples are included. Section 3 applies the theory to some realistic situations. Section 4 concludes the paper with a few concluding remarks. The proofs of all our stated results are available upon request.

2. A Dynamic Queuing Model

Suppose that a supplier provides a product for distribution among consumers. For some reason, the market price fails to adjust according to the demand or simply there is no market for the product. We can thus simply assume a fixed price for the product (zero price if there is no market). Under this situation, when there is excess demand, there will be a queue for the product, unless the supplier takes measures to prevent the queue.

As is the case in any queuing model, suppose that there is a capacity constraint on the supplier so that the supplier can only provide the product to a limited number of consumers. However, different from a traditional model, the supplier in our model is willing to provide multiple shifts, meaning that when a group of consumers finish consumption, another group of consumers will be invited to consume, and so on.

2.1. The Setup

As explained above, the firm offers n shifts with a constant price p for all the shifts and with a fixed capacity k that limits the quantity of the product supplied in each shift. There is a continuum of consumers who are evenly distributed on a set \mathbb{X} , where $\mathbb{X} \subset \mathbb{R}_+$.²

A typical consumer is indexed by a location $x \in \mathbb{X}$. Each consumer has a unit demand for the product, consuming either one unit or none. The consumers differ from each other in three aspects: the **valuation** $v(x) \geq 0$ of the product, the **wage rate** $w(x) > 0$, and the **discount factor** $\delta(x) > 0$ of time preferences. All of the three variables are denominated in dollars. Call $v(\cdot)$, $w(\cdot)$, and $\delta(\cdot)$ the distributions of valuations, wage rates, and the discount factors, respectively. Thus, if a consumer consumes the product in the i th shift, queues for t_i units of time and pays p for the product, the consumer's surplus is $v(x) - p - t_i w(x) - (i - 1)\delta(x)$, where $t_i w(x)$ is the monetary cost of queuing and $(i - 1)\delta(x)$ is the monetary cost of delay.

Denote $\hat{v}(x)$ as the consumer's valuation of the product in time units and $\hat{\delta}(x)$ as the discount factor in time units, i.e.,

$$\hat{v}(x) \equiv \frac{v(x) - p}{w(x)}, \quad \hat{\delta}(x) \equiv \frac{\delta(x)}{w(x)}.$$

Call $\hat{v}(x)$ the consumer's **time valuation of the product** and $\hat{\delta}(x)$ the **time discount factor**, and call $\hat{v}(\cdot)$ and $\hat{\delta}(\cdot)$ the distributions of time valuations and time discount factors, respectively.

Each consumer is assumed to know the price, the number of shifts, and his own valuation, costs and discount rate. But he does not need to know the capacity k and other consumers' valuation, wage rates and discount factors.

²For the users of the service, the number of shifts n , the price p and the capacity k are taken as given. These variables are determined by the provider of the service. As typical, the provider's problem is not discussed in a queuing model. See, for example, an application of this model in Wang–Zhu (2003), where the consumers face different dynamic queuing problems in different seasons while the firm chooses different numbers of shifts in different seasons, different prices in different seasons, and a capacity for all seasons.

Since the consumers discount delayed consumption, they all would like to consume in as early a shift as possible. Consumption in early shifts is rationed by the waiting-line competition in which the first-come-first-served rule determines the sequence of consumption. In addition, the consumers observe the following two rules of behavior.

Rule 1: If a consumer is indifferent between any two shifts, he will consume in the earlier shift.

Rule 2: All consumers follow the “get-it-while-you-can” strategy when they compete for the product in queues.

Rule 1 is made to avoid the tie situation. Rule 2 avoids the situation in which a consumer does not see the need to join a queue if no one has joined the queue; but once one person joins the queue, everyone rushes in [see Bagnoli *et al.* (1989) for a detailed discussion of Rule 2].

Before the waiting-line competition begins, and taking as given the price, the capacity and the number of shifts, the consumers form expectations about the length of queuing time for each shift. With these expectations, each of the consumers first decides whether he will consume or not; if he decides to consume, then he must decide in which shift he should compete in a queue in order to maximize his surplus. In equilibrium, the expected queuing length equals the actual queuing length in every shift.

2.2. The Optimization Problem

For a consumer x , his problem is: given the capacity k , the number of shifts n , and the price p , he forms expectations of the lengths of queuing time t_1, t_2, \dots, t_n with $t_i \geq 0$ and considers a dynamic queuing problem:

$$\begin{aligned} \max_{1 \leq i \leq n} \quad & v(x) - p - t_i w(x) - (i - 1)\delta(x) \\ \text{s.t.} \quad & v(x) - p \geq t_i w(x) + (i - 1)\delta(x), \end{aligned}$$

or, equivalently,

$$\begin{aligned} \min_{1 \leq i \leq n} \quad & t_i + (i - 1)\hat{\delta}(x) \\ \text{s.t.} \quad & \hat{v}(x) \geq t_i + (i - 1)\hat{\delta}(x). \end{aligned} \tag{1}$$

We will now try to simplify this problem. Problem (1) implies that consumer x will choose to consume in shift i if

$$IC: \quad t_j + (j - 1)\hat{\delta}(x) > t_i + (i - 1)\hat{\delta}(x), \quad \text{for all } j \neq i,$$

and

$$IR: \quad \hat{v}(x) \geq t_i + (i - 1)\hat{\delta}(x).$$

The first inequality is the **incentive compatibility (IC)** condition and the second is the **individual rationality (IR)** condition. The IC condition is the same as an IC condition in agency models; since the supplier cannot force an individual to consume in a certain shift, the condition induces the individual to choose a certain shift. The IR condition determines whether or not a consumer is interested in the product at all, given the costs.

Define \mathcal{A}_i and \mathcal{B}_i , respectively, as

$$\begin{aligned} \mathcal{A}_i &\equiv \left\{ x \geq 0 \mid t_j + j\hat{\delta}(x) > t_i + i\hat{\delta}(x), \forall j \neq i \right\}, \\ \mathcal{B}_i &\equiv \left\{ x \geq 0 \mid \hat{v}(x) \geq t_i + (i - 1)\hat{\delta}(x) \right\}. \end{aligned}$$

Indeed, \mathcal{A}_i is the set of consumers who satisfy the IC constraint for shift i , and \mathcal{B}_i is the set of consumers who satisfy the IR constraint for shift i . Also, denote \mathcal{B} as the set of consumers who will consume (in some shift):

$$\mathcal{B} \equiv \left\{ x \geq 0 \mid \exists j, j \leq n, \text{ s.t. } \hat{v}(x) \geq t_j + (j - 1)\hat{\delta}(x) \right\}.$$

We have $\mathcal{B} = \bigcup_{i=1}^n \mathcal{B}_i$, and the following lemma is intuitive.

Lemma 1. *Problem (1) is equivalent to the following problem*

$$\begin{aligned} \min_{1 \leq i \leq n} \quad & t_i + (i - 1)\hat{\delta}(x) \\ \text{s.t.} \quad & x \in \mathcal{B}. \end{aligned} \tag{2}$$

2.3. The Queuing Equilibrium

Given a sequence of expected queuing lengths t_1, \dots, t_n (formed ex ante), each individual solves his own problem in (2). The solutions from

all the consumers constitute the demand in each shift. With the capacity constraint in each shift, an ex-post queuing length t_i^* is determined for each shift i . When the ex-ante queuing length is the same as the ex-post queuing length for each shift, we have an equilibrium.

Definition 1. A rational expectations equilibrium (REE) of queuing (t_1^*, \dots, t_n^*) is realized if the ex post queuing time in each of the shifts is the same as the ex ante queuing time. That is, $t_i^* = t_i$ for all i . ■

Our definition of REE differs from equilibrium queuing concepts in traditional models. In our model, individuals form expectations on queuing lengths. When their expectations are consistent with the actual queuing lengths, an equilibrium is reached. Individuals in our model do not need to know other players' valuations, wage rates and discounts. Traditional queuing models are often based the Bayesian equilibrium concept in which each individual knows the distribution functions of all his opponents and forms expectations on the opponents' willingness to queue. By this, each player can estimate the equilibrium queuing time and decide whether or not to queue. In this case, a symmetric equilibrium is usually adopted.³

For any two shifts i and j , $i < j$, if

$$t_i + (i - 1)\hat{\delta}(x) = t_j + (j - 1)\hat{\delta}(x), \quad \text{or} \quad \hat{\delta}(x) = -\frac{t_i - t_j}{i - j},$$

then consumer x is indifferent regarding the two shifts. A shift is said to be **strictly most preferred** for consumer x if the shift is strictly preferred by him over all the other shifts.

Lemma 2. Denote $t_0 \equiv +\infty$ and $t_{n+1} \equiv t_n$. For any REE equilibrium (t_1^*, \dots, t_n^*) ,

³There are three disadvantages of the Bayesian approach:

1. It can only deal with finite players.
2. Each player needs to know a lot of information.
3. A set of distribution functions on players' valuations needs to be introduced. And, the equilibrium is dependent on these distribution functions.

One disadvantage of our REE is that it may not be efficient (as defined later). However, we will show the existence of a unique efficient REE.

- (a) t_i^* is decreasing across shifts: $t_1^* > \dots > t_n^* \geq 0$.
- (b) t_i^* is decreasing at a decreasing rate: $t_{i-1}^* - t_i^* \geq t_i^* - t_{i+1}^*$, for all i .
- (c) The set of consumers who strictly most prefer shift i is

$$\left\{ x \geq 0 \mid \max_{i < j \leq n} \left| \frac{t_j^* - t_i^*}{j - i} \right| < \hat{\delta}(x) < \min_{1 \leq j < i} \left| \frac{t_j^* - t_i^*}{j - i} \right| \right\}. \blacksquare$$

Loosely speaking, Lemma 2(c) says that if a consumer strictly most prefers shift i , then those consumers whose types are similar to him will also choose shift i .

Theorem 1. (Characterization of Queuing Equilibria). For any rational expectations equilibrium $\{t_i^*\}_{i=1}^n$, there exists a convex and decreasing function $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $\phi(i) = t_i^*$ for all i . \blacksquare

Figure 1 depicts the queuing function $\phi(\cdot)$ in Theorem 1.

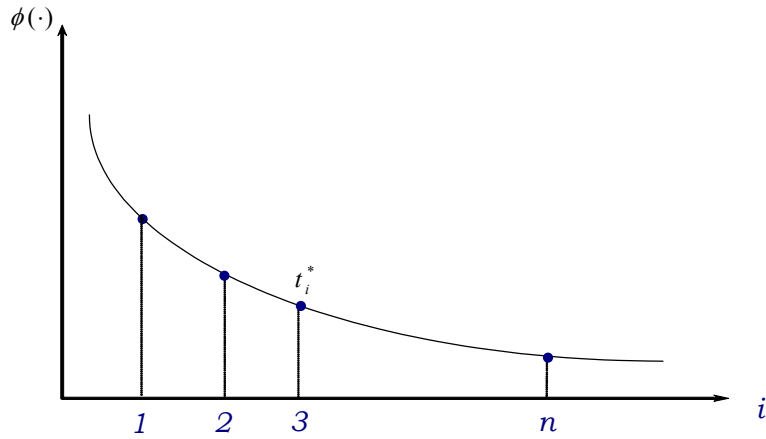


Figure 1. Convex and Decreasing Queuing Length Across Shifts

Lemma 3. Denote $t_0 \equiv +\infty$ and $t_{n+1} \equiv t_n$. Under Rules 1 and 2, for any shift i , we have

$$\mathcal{A}_i = \hat{\delta}^{-1}[t_i^* - t_{i+1}^*, t_{i-1}^* - t_i^*]. \blacksquare$$

Lemma 3 says that consumer x satisfies the *IC* condition for shift i if and only if his/her time valuation $\hat{\delta}(x)$ is in the interval $[t_i^* - t_{i+1}^*, t_{i-1}^* - t_i^*]$. That is, a consumer x who picks shift i satisfies the conditions $t_i^* + \hat{\delta}(x) < t_{i-1}^*$ and $t_i^* < t_{i+1}^* + \hat{\delta}(x)$, meaning that the queuing time plus delay cost for shift i is less than the queuing time for the earlier shift, and the queuing time for shift i is less than the queuing time plus delay cost for the later shift. This result is very convenient, since the decision on shift i is related only to its neighboring two shifts.

By Lemmas 1–3, we have reduced problem (1) to a simple form so that we can now easily define the effective demand. Denote $\mathcal{D}_i \equiv \mathcal{A}_i \cap \mathcal{B}_i$ and call it the **demand set** of shift i . \mathcal{D}_i contains all the consumers who consume in shift i . By Lemma 3, we have

$$\mathcal{D}_i = \{x \geq 0 \mid t_i - t_{i+1} \leq \hat{\delta}(x) < t_{i-1} - t_i, t_i + (i-1)\hat{\delta}(x) \leq \hat{v}(x)\}, \quad (3)$$

for $i = 1, \dots, n$, where we have arbitrarily set $t_0 \equiv +\infty$ and $t_{n+1} \equiv t_n$ for simplicity of notation. Further, denote $\mathcal{D} \equiv \mathcal{B}$ and call it the **total demand set**. \mathcal{D} contains all the consumers who will consume. By Lemma 1, we have

$$\mathcal{D} = \bigcup_{i=1}^n \mathcal{D}_i.$$

Let $\mu(\cdot)$ be the Lebesgue measure on \mathbb{R} .⁴ Since $\hat{\delta}$ and \hat{v} are continuous, \mathcal{D}_i and \mathcal{D} are Borel-measurable. Now, define $x^d \equiv \mu(\mathcal{D})$ and call it the **effective demand**. In contrast, we will call the consumer's valuation function, $v(\cdot)$, the (inverse) **ordinary demand**. The effective demand x^d differs from the ordinary demand because it includes the costs of queuing and consumption delay. Consequently, the effective demand depends not only on price, p , but also on capacity, k . In **equilibrium**, it must be the case that there exists n^* , $1 \leq n^* \leq n$, such that $\mu(\mathcal{D}_i) = k$ for $i \leq n^* - 1$, and $\mu(\mathcal{D}_{n^*}) \leq k$.

⁴The assumption of an even distribution of the consumers on $\mathbb{X} \subset \mathbb{R}_+$ is made without loss of generality. In general, if the distribution of consumers is described by an accumulative distribution function $F(x)$, we should replace the Lebesgue measure $\mu(\cdot)$ by the Riemann-Stieltjes measure $\mu_F(\cdot)$, where for any subset $A \subset \mathbb{X}$ we have $\mu_F(A) = \int_A dF(x)$.

Theorem 2. (Existence of REE). Suppose $\hat{v}(\cdot)$ is essentially bounded from above. Given any price $p \geq 0$, capacity $k > 0$ and n shifts, if the distribution of the time valuations $\hat{v}(\cdot)$ and the distribution of the time discount factors $\hat{\delta}(\cdot)$ are continuous, then there exists a rational expectations equilibrium $(t_1^*, t_2^*, \dots, t_n^*)$, $t_1^* \geq t_2^* \geq \dots \geq t_n^* \geq 0$.⁵ ■

One key advantage of an rational expectations equilibrium is that each consumer does not need to know the preferences and income levels of other consumers. Every consumer forms his own expected queuing lengths, from which he makes a decision based on his own preferences and income level. When all these decisions balance out in aggregate, we have a rational expectations equilibrium. This approach departs from the traditional approach in queuing theory, where every consumer has the knowledge of the joint distribution function of other consumers' preferences and income levels; see, for example, Taylor–Tsui–Zhu (2003). In the traditional theory, this distribution function matters to the solution, while in a rational expectations equilibrium, there is no such a distribution function. Our approach is crucial in reducing the complexity of a multi-period queuing problem.

Generally speaking, the rational expectations equilibrium is not unique. To see this, consider the following example.

Example 1. (Multi-Shift Queuing Model). Consider the case when the demand is linear. Specifically, let

$$v(x) = 1 - x, \quad w(x) = \beta v(x), \quad \delta(x) = \delta_0, \quad \text{for } x \in \mathbb{X} = [0, 1],$$

where β and δ_0 are constants with $\beta > 0$ and $\delta_0 \geq 0$. Let $n = 2$. Further, for convenience, assume $k \geq \frac{1}{2}$ so that $\mu(\mathcal{D}_2) \leq k$, by which $t_2^* = 0$. Denote $t = t_1$. By (3), we have

$$\mathcal{D}_1 = \{x \mid tw(x) \leq \delta(x), \quad tw(x) + p \leq v(x)\}, \quad (4a)$$

$$\mathcal{D}_2 = \{x \mid tw(x) > \delta(x), \quad \delta(x) + p \leq v(x)\}, \quad (4b)$$

⁵The definition of **essential boundedness** can be found in Lang (1993). The boundedness of \hat{v} rules out the possibility of an infinite queuing length. Notice that v , w and δ need to be continuous but are not required to be monotonic. We can actually allow the functions to be piecewise continuous. If so, they can be step functions, which means that our model includes the case of finite consumers as a special case.

implying

$$\begin{aligned}\mathcal{D}_1 &= \{x \mid \beta t(1-x) \leq \delta_0, \quad \beta t(1-x) + p \leq 1-x\}, \\ \mathcal{D}_2 &= \{x \mid \beta t(1-x) > \delta_0, \quad \delta_0 + p \leq 1-x\}.\end{aligned}$$

We must have $t < \frac{1}{\beta}$, otherwise $\beta t(1-x) \geq 1-x \geq 1-x-p$, implying $\mu(\mathcal{D}_1) = 0$, which cannot be an equilibrium situation.⁶ Thus, $1 - \beta t > 0$ and

$$\begin{aligned}\mathcal{D}_1 &= \left\{ x \mid x \geq 1 - \frac{\delta_0}{\beta t}, \quad x \leq 1 - \frac{p}{1 - \beta t} \right\}, \\ \mathcal{D}_2 &= \left\{ x \mid x < 1 - \frac{\delta_0}{\beta t}, \quad x \leq 1 - \delta_0 - p \right\}.\end{aligned}$$

We also must have

$$1 - \frac{\delta_0}{\beta t} < 1 - \frac{p}{1 - \beta t}, \quad \text{or} \quad t < \frac{\delta_0}{\beta(p + \delta_0)}, \quad (5)$$

otherwise $\mu(\mathcal{D}_1) = 0$. By condition (5), we have

$$\mathcal{D}_1 = \left[1 - \frac{\delta_0}{\beta t}, \quad 1 - \frac{p}{1 - \beta t} \right].$$

Also, by condition (5), we have $1 - \frac{\delta_0}{\beta t} < 1 - \delta_0 - p$, implying

$$\mathcal{D}_2 = \left[0, \quad 1 - \frac{\delta_0}{\beta t} \right).$$

⁶Condition $t < \frac{1}{\beta}$ simply says that the queuing cannot be too long otherwise no one wants to be in the first shift. However, if no one is in the first shift, there would be no queue and then everyone wants to be in the first shift. Such a situation cannot possibly be in equilibrium.

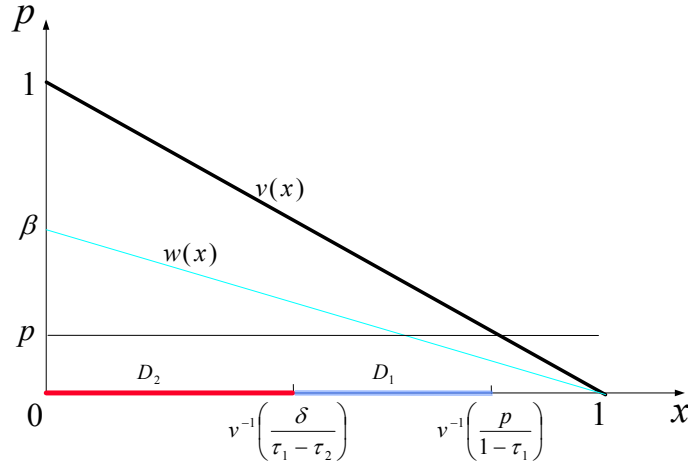


Figure 2. Two-Shift Queuing Equilibria

The equilibrium queuing time t^* is determined by the capacity constraint in the first shift:

$$k = \left(1 - \frac{p}{1 - \beta t}\right) - \left(1 - \frac{\delta_0}{\beta t}\right),$$

which gives two solutions

$$t^* = \frac{1}{2\beta k} \left[p + \delta_0 + k \pm \sqrt{(p + \delta_0 + k)^2 - 4\delta_0 k} \right]. \quad (6)$$

Both queuing times are positive and thus valid, but one is longer than the other. Clearly, the longer queuing time is less efficient. This leads to our definition of an efficient REE in Definition 2. ■

Definition 2. An **efficient rational expectations equilibrium (EREE) of queuing** is a rational expectations equilibrium for which *there is no other equilibrium that offers a shorter queuing time in any shift.* ■

Theorem 3. (Existence and Uniqueness of EREE). Suppose $\hat{v}(\cdot)$ is essentially bounded from above. For any given price $p \geq 0$, capacity $k > 0$ and number of shifts n , if the distribution of time valuations $\hat{v}(\cdot)$ and the distribution of time discount factors $\hat{\delta}(\cdot)$ are continuous, then there exists a unique efficient rational expectations equilibrium $(t_1^*, t_2^*, \dots, t_n^*)$, $t_1^* \geq t_2^* \geq \dots \geq t_n^* \geq 0$. ■

The EREE is a rational expectations equilibrium that is the most efficient within our queuing model and is unique by definition. The two notions of equilibria can be distinguished by their stability characteristics: the EREE is stable, while other queuing equilibria are not.

A comparison between our multi-shift queuing model with the traditional single-shift queuing model is interesting. The following example is for this purpose.

Example 2. (Single-Shift Queuing Model). In the traditional model, multiple shifts are not allowed, i.e., $n = 1$. For consumer x , his problem is: given capacity k and price p , he forms expectations of the length of queuing time t , with $t \geq 0$, and decides to participate if and only if $v(x) \geq p + tw(x)$. Thus,

$$\mathcal{D}_1 = \mathcal{B} = \mathcal{B}_1 = \{x \geq 0 \mid v(x) \geq p + tw(x)\}.$$

In the case of linear demand in Example 1,

$$\mathcal{D}_1 = \left[0, 1 - \frac{p}{1 - \beta t}\right].$$

The equilibrium condition $\mu(\mathcal{D}_1) = k$ implies

$$\bar{t} = \frac{1}{\beta} \left(1 - \frac{p}{1 - k}\right). \quad (7)$$

As expected, by comparing the efficient solution in (6) with the solution in (7), we find $\bar{t} > t^*$ for any parameter values. ■

How costly the queuing mechanism is in terms of social welfare is also interesting. In the following example, we show that the multi-shift queuing equilibrium implies only a small loss of social welfare in some cases and that allowing more shifts can dramatically reduce the loss.

Example 3. (Welfare Loss). Suppose the firm produces the product at a constant marginal cost $c \geq 0$. Using the functions chosen in Example 1, for the efficient queuing equilibrium t^* in Example 1, the social welfare is

$$\begin{aligned} W^* &= \int_0^k [v(x) - t^*w(x)]dx + \int_k^{1 - \frac{p}{1 - \beta t^*}} [v(x) - \delta_0]dx - c \left(1 - \frac{p}{1 - \beta t^*}\right) \\ &= \frac{1}{2} \left[(1 - c - \delta_0)^2 + 2(\delta_0 - \beta t^*)k + \beta t^*k^2 - \left(\frac{p}{1 - \beta t^*} - c - \delta_0\right)^2 \right]. \end{aligned}$$

For the queuing equilibrium \bar{t} in Example 2, the social welfare is

$$\bar{W} = \int_0^k [v(x) - \bar{t}w(x)]dx - ck = \frac{pk}{2} \frac{2-k}{1-k} - ck.$$

For the price mechanism, the prices in both shifts are allowed to adjust to maximize social welfare. The social welfare maximization problem is

$$W^{**} \equiv \max_{p_1, p_2 \geq 0, v^{-1}(p_1) \leq k} \int_0^{v^{-1}(p_1)} [v(x) - t^{**}w(x)]dx + \int_{v^{-1}(p_1)}^{v^{-1}(p_2)} [v(x) - \delta_0]dx - cv^{-1}(p_2),$$

where p_1 and p_2 are the prices in the first and second shifts, respectively, and t^{**} is the equilibrium queuing time if the market in the first shift is not cleared. Obviously, the optimal price p_1^* in the first shift must be the one that clears the market (eliminating the loss from queuing), i.e., $v^{-1}(p_1^*) = k$. Thus, the problem becomes

$$W^{**} \equiv \max_{p_2 \geq 0} \int_0^{v^{-1}(p_2)} v(x)dx - (c + \delta_0)v^{-1}(p_2) + \delta_0k,$$

implying $p_2^{**} = c + \delta_0$ and

$$W^{**} = \frac{1}{2}(1 - c - \delta_0)^2 + \delta_0k.$$

To have some idea on the magnitude of the losses, we try out some parameter values. We first arbitrarily assign $c = 0.1$, $p = 0.2$, $\beta = 1$ and $k = 0.6$. We find that these four parameters are not important in evaluating the losses (besides the requirement of $0.5 \leq k \leq 1$ and a small differential $p - c$). However, δ_0 is important. We find

$$\frac{W^{**} - W^*}{W^{**}} = \begin{cases} 13.5\% & \text{if } \delta_0 = 0.1, \\ 7.1\% & \text{if } \delta_0 = 0.05, \\ 2.4\% & \text{if } \delta_0 = 0.01. \end{cases}$$

In other words, the loss from queuing can be large if people hate waiting, but it can be very small if people do not mind waiting. We also find

$$\frac{W^{**} - \bar{W}}{W^{**}} = \begin{cases} 60.5\% & \text{if } \delta_0 = 0.1, \\ 61.7\% & \text{if } \delta_0 = 0.05, \\ 62.7\% & \text{if } \delta_0 = 0.01. \end{cases}$$

In other words, by allowing multiple shifts, even with only one more shift, the welfare improvement is huge. If more shifts are allowed, the loss from queuing should be much smaller.

Queues exist everywhere in our lives even though the capitalist system is believed to have become quite complete and perfect. This example shows that a multi-shift queuing system can be fairly efficient if the delay cost is small for many consumers. This may explain the continuing existence of queuing. There have been extensive studies on pricing a queuing system in order to improve social welfare.⁷ However, proposed pricing systems generally involve some social costs (e.g., administrative costs), which may thus not necessarily be welfare improving in comparison with a multi-shift queuing system without an administrative cost. ■

3. Applications

There are many possible applications of our dynamic queuing model. To see different characteristics of some applications, we consider three distinct cases: shopping, highways, and restaurants. Again, for simplicity, we restrict the number of shifts to two and we will assume that the consumers are located in $[0, 1]$ and $k \geq \frac{1}{2}$.

3.1. Shopping

Some shoppers are more queue averse than are others. It is generally observed that low-income people are likely to stand in a long queue. In other words, given a product that has little difference in valuation and in

⁷See, for example, Leeman (1964), Naor (1969), Nichols–Smolensky–Tideman (1971), Stahl–Alexeev (1985), Alexeev (1989), and Polterovich (1993). Externality is the basic argument in this line of research.

aversion towards delay in consumption among people, the perceived cost of queuing can differ greatly among people. To highlight this point, we assume a changing wage function (perceived penalty for queuing) but assume constants for the other two functions. In particular, we assume

$$\begin{aligned} v(x) &= v_0, \\ w(x) &= 1 - x, \\ \delta(x) &= \delta_0, \end{aligned}$$

where $v_0 > 0$ and $\delta_0 > 0$ are two constants. In this case, (4a)-(4b) become

$$\begin{aligned} \mathcal{D}_1 &= \left\{ x \mid x \geq 1 - \frac{\delta_0}{t}, \quad x \geq 1 - \frac{v_0 - p}{t} \right\}, \\ \mathcal{D}_2 &= \left\{ x \mid x < 1 - \frac{\delta_0}{t}, \quad \delta_0 + p \leq v_0 \right\}. \end{aligned}$$

If $\delta_0 + p > v_0$, no one will consume in the second shift. Also, in this case, we have

$$1 - \frac{\delta_0}{t} < 1 - \frac{v_0 - p}{t},$$

implying

$$\mathcal{D}_1 = \left[1 - \frac{v_0 - p}{t}, 1 \right], \quad \mathcal{D}_2 = \emptyset.$$

The equilibrium condition $\mu(\mathcal{D}_1) = k$ implies

$$t^* = \frac{v_0 - p}{k}.$$

If $\delta_0 + p \leq v_0$, then

$$\mathcal{D}_1 = \left[1 - \frac{\delta_0}{t}, 1 \right], \quad \mathcal{D}_2 = \left[0, 1 - \frac{\delta_0}{t} \right).$$

The equilibrium condition $\mu(\mathcal{D}_1) = k$ then implies

$$t^* = \frac{\delta_0}{k}.$$

In both cases, low-income people will queue while high-income people will not. In the first case, since the price and/or the cost of delay are too high relative to valuation, there is no demand in the second shift. In the second case, the costs are justified by the valuation and everyone will consume; some prefer to queue for early consumption, while others avoid queuing by late consumption.

3.2. Highways

For the use of highways, being stranded on a road can have major consequences for some but not so for others. For example, for those who drive to work, it can be very costly to be stranded in traffic; but for those who go shopping, being stranded is just a minor irritation. So, the valuation of highways in the morning rush hour is very different among different people. To highlight this point, we assume a changing valuation function, but assume constants for the other two functions. In particular, we assume

$$\begin{aligned} v(x) &= 1 - x, \\ w(x) &= w_0, \\ \delta(x) &= \delta_0, \end{aligned}$$

where $w_0 > 0$ and $\delta_0 > 0$ are two constants. In this case, (4a)-(4b) become

$$\begin{aligned} \mathcal{D}_1 &= \{x \mid tw_0 \leq \delta_0, \quad tw_0 + p \leq 1 - x\}, \\ \mathcal{D}_2 &= \{x \mid tw_0 > \delta_0, \quad \delta_0 + p \leq 1 - x\}. \end{aligned}$$

We must have

$$tw_0 \leq \delta_0, \tag{8}$$

otherwise $\mathcal{D}_1 = \emptyset$, which cannot be an equilibrium situation, as explained in Footnote 6. Then,

$$\begin{aligned} \mathcal{D}_1 &= \{x \mid tw_0 + p \leq 1 - x\} = [0, 1 - p - tw_0], \\ \mathcal{D}_2 &= \emptyset. \end{aligned}$$

The equilibrium condition $\mu(\mathcal{D}_1) = k$ then implies

$$t^* = \frac{1 - p - k}{w_0}.$$

With this solution, condition (8) becomes a restriction on the parameters: $p + k + \delta_0 \geq 1$.

The solution indicates that only the high-valuation people will be on the road in the morning rush hour. Other people, such as shoppers, will wait until after the rush hour (when there is no queue on the road).

3.3. Restaurants

In going to restaurants, the timing is an important consideration. For working people, lunch hours are fixed by companies and workers must have lunch on time, while others, such as shoppers, may not mind to have some delay. To highlight this point, assume a changing discount factor, but assume constants for the other two functions. In particular, we assume

$$\begin{aligned}v(x) &= v_0, \\w(x) &= w_0, \\ \delta(x) &= 1 - x,\end{aligned}$$

where $v_0 > 0$ and $w_0 > 0$ are two constants. In this case, (4a)-(4b) become

$$\begin{aligned}\mathcal{D}_1 &= \{x \mid tw_0 \leq 1 - x, \quad tw_0 + p \leq v_0\}, \\ \mathcal{D}_2 &= \{x \mid tw_0 > 1 - x, \quad 1 - x + p \leq v_0\}.\end{aligned}$$

First of all, we must have

$$tw_0 + p \leq v_0, \tag{9}$$

otherwise $\mathcal{D}_1 = \emptyset$, which cannot be an equilibrium situation. With (9), we have

$$\mathcal{D}_1 = [0, 1 - tw_0].$$

Given (9), we have $1 - tw_0 \geq 1 + p - v_0$, implying

$$\mathcal{D}_2 = \{x \mid x > 1 - tw_0, \quad x \geq 1 + p - v_0\} = (1 - tw_0, 1].$$

The equilibrium condition $\mu(\mathcal{D}_1) = k$ then implies

$$t^* = \frac{1 - k}{w_0}.$$

With this solution, condition (9) becomes a restriction on the parameters: $1 + p \leq k + v_0$.

The solution indicates that everyone will have lunch, and those who must eat at lunch time will queue for early consumption, while others will wait until after the lunch hour.

4. Concluding Remarks

For many goods, demand fluctuates greatly over different seasons and across different time frames. In a long-term plan, a firm/government will set up an appropriate capacity that balances the demands across seasons. Since the capacity is costly to set up and to maintain, the firm usually has a binding capacity during a period of high demand. In many realistic situations, the price is sticky, resulting in queues for the product. Instead of a single-shot game of queuing, we consider a multi-shift game of queuing. We believe that our dynamic queuing model fits many realistic situations better than do many existing models.

Multiple shifts reduce the cost of queuing and allow consumers with different preferences to choose different consumption strategies. Also, the firm can save the cost of capacity by repeatedly using its capacity. In fact, if there is a cost for price changes, our multi-shift queuing mechanism can be more efficient than the price mechanism.

There is extensive research in the literature on multi-server queuing models. Knudsen (1972) is the pioneer. In these models, the service provider provides several service lines/points at the same time, instead of a single server. Multi-server models are initially proposed as a way to reduce the social cost of queuing, just as what Example 3 has shown for our model. These models are complementary to our model. In multi-server models, the attitude about delay does not play a role, while this feature is crucial in our analysis, as shown in Example 3. These two types of models fit different types of applications, and, in many applications, both features appear. An additional server means additional costs, especially the cost of additional capacity, while an additional shift means more extensive use of the existing capacity. In some cases, such as highways and subways, since more servers cannot justify the extra cost, more shifts are used, while in other cases, such as buses and elevators, more servers are used. However, in many cases, such as buses, restaurants and airlines, both multiple shifts and multiple servers are used by companies/governments as a way to increase profit/welfare.

Queuing is everywhere in life. It is much more popular than any other nonprice mechanism. From the moment you get out of bed, you may have to queue for using a bathroom, queue for breakfast, queue for using a road or to get on a bus, queue for using an elevator, queue for seeing a doctor, queue to get cash from a bank, and so on, until you are finally back home. Queuing is also everywhere in research. It is an important research subject in

departments of economics, management science, statistics, and operations research. There are several reasons for the popularity of queuing. First, it can exist without an administrative cost. A queue appears and disappears automatically by itself depending on the needs. For a small scale activity such as the use of an elevator, the administrative cost of using a price mechanism or a lottery can be very large in comparison with its scale of operation. Second, there is no risk on whether and when you will have your turn for consumption. You choose a queue in a certain shift and you are guaranteed to consume at certain point of time. In a lottery, however, you face the risk of not being chosen before your deadline. This uncertainty can be very costly for some. Third, with multiple shifts, queuing can be nearly as efficient as the price mechanism. We therefore believe that a multi-shift queuing model can contribute a lot to research on queuing.

References

- [1] Alexeev, M. (1989): “A Note on Privileges in a Queue-Rationed CPE with Black Markets”, *Journal of Economic Theory*, 47, 422–430.
- [2] Bagnoli, M., S.W. Salant, and J.E. Swierzbinski (1989): “Durable-Goods Monopoly with Discrete Demand”, *Journal of Political Economy*, 97, 1459-1478.
- [3] Bennett, J. (1991): “Repressed Inflation, Queuing and the Resale of Goods in a Centrally Planned Economy”, *European Economic Review*, 35 (1), 49–60.
- [4] Barzel, Y. (1974): “A Theory of Rationing by Waiting”, *Journal of Law and Economics*, 17, 73–95.
- [5] Daniel, J. (1995): “Congestion Pricing and Capacity of Large Hub Airports: A Bottleneck Model with Stochastic Queues”, *Econometrica*, 63, 327–367.
- [6] Davidson, C. (1988): “Equilibrium in Servicing Industries: An Economic Application of Queuing Theory”, *Journal of Business*, 61 (3), 347–367.
- [7] Deacon, R.T. and J. Sonstelie (1985): “Rationing by Waiting and the Value of Time: Results from a Natural Experiment”, *Journal of Political Economy*, 93, 637–647.

-
- [8] DeVany, A. (1976): “Uncertainty, Waiting Time, and Capacity Utilization: A Stochastic Theory of Product Quality”, *Journal of Political Economy*, 84, 523–540.
- [9] Dolan, R.J. (1978): “Incentive Mechanisms for Priority Queuing Problems”, *Bell Journal of Economics*, 9, 421–436.
- [10] Hassin, R. and M. Haviv (2003): *To Queue or Not to Queue: Equilibrium Behavior in Queuing Systems*, London: Kluwer Academic Publisher.
- [11] Holt, C. and R. Sherman (1982): “Waiting Line Auction”, *Journal of Political Economy*, 90, 280–294.
- [12] Knudsen, N.C. (1972): “Individual and Social Optimization in a Multi-Server Queue with a General Cost-Benefit Structure”, *Econometrica*, 40, 515–528.
- [13] Lang, S. (1993): *Real and Functional Analysis*, 3rd ed., Springer-Verlag.
- [14] Leeman, W.A. (1964): “The Reduction of Queues Through the Use of Price”, *Operation Research*, 12, 783–785.
- [15] Lui, F.T. (1985): “An Equilibrium Queuing Model of Bribery”, *Journal of Political Economy*, 93, 760–781.
- [16] Naor, P. (1969): “The Regulation of Queue Size by Levying Tolls”, *Econometrica*, 37, 15–24.
- [17] Nichols, D., E. Smolensky and T. Tideman (1971): “Discrimination by Waiting Time in Merit Goods”, *American Economic Review*, 61, 312–323.
- [18] Polterovich, V. (1993): “Rationing, Queues, and Black Markets”, *Econometrica*, 61, 1–28.
- [19] Sattinger, M. (2002): “A Queuing Model of the Market for Access to Trading Partners”, *International Economic Review*, 43 (2), 533–547.
- [20] Shi, Shuzhong. (1990): *Convex Analysis*, Shanghai Science and Technology Publishing Company.

-
- [21] Stahl, D.O. and M. Alexeev (1985): “The Influence of Black Markets on a Queue-Rationed Centrally Planned Economy”, *Journal of Economic Theory*, 35, 234–250.
- [22] Suen, W. (1989): “Rationing and Rent Dissipation in the Presence of Heterogeneous Individuals”, *Journal of Political Economy*, 97, 1384–1394.
- [23] Taylor, G.A., K. Tsui, and L. Zhu (2003): “Lottery or Waiting-Line Auction?” *Journal of Public Economics*, 87, 1313–1334.
- [24] Vickrey, W.S. (1969): “Congestion Theory and Transport Investment”, *American Economic Review*, 59, 251–260.
- [25] Wang, S. and L. Zhu (2003): “Variable Capacity Utilization and Countercyclical Pricing”, Working Paper, HKUST.