Queueing Syst (2017) 86:141–212
DOI 10.1007/s11134-017-9521-6

# An equilibrium analysis of a discrete-time Markovian queue with endogenous abandonments

**Baris Ata[1] · Peter W. Glynn[2] · Xiaoshan Peng[1]**

Received: 26 September 2015 / Revised: 23 February 2017 / Published online: 3 April 2017


**Abstract** This paper studies a Geo/Geo/1+$GI$ queue in which the abandonments are endogenous. One crucial feature of this model is that the abandonment behavior is affected by the system performance and vice versa. Our model captures this interaction by developing two closely related models: an abandonment model and a queueing model. In the abandonment model, customers take the virtual waiting time distribution as given. They receive a reward $r$ from service and incur a cost $c$ per period of waiting. Customers are forward-looking and maximize their expected discounted utilities by making wait or abandon decisions dynamically as they wait in the queue. The queueing model takes the customers' abandonment time distribution as an input and studies the resulting virtual waiting time distribution. In equilibrium, the customers' abandonment behavior and the system performance must be consistent across the two models. Therefore, combining the two models and imposing this consistency requirement, we show that there exists a unique equilibrium. Lastly, we provide a computational scheme to calculate the equilibrium numerically.

**Keywords** Invisible queue · Endogenous abandonment · Customer abandonment · Impatient customers · Single-class queue

✉ Xiaoshan Peng
x-peng@chicagobooth.edu

Baris Ata
Baris.Ata@chicagobooth.edu

Peter W. Glynn
glynn@stanford.edu

1 5807 South Woodlawn Avenue, Chicago, IL 60637, USA

2 Huang Engineering Center 357, Stanford, CA 94305, USA

**Mathematics Subject Classification** 60K25 · 90B22 · 91G80

## 1 Introduction

Queueing models with abandonments have been studied extensively in the operations research literature both from performance analysis and from optimization perspectives; see [40] for a survey. The traditional approach endows customers with exogenous patience time distributions, and a customer abandons when her waiting time exceeds her patience time (abandonment time). Exact results are rare and require distributional assumptions on the arrival, service and abandonment processes. Examples include [15, 20,22,38]; see [2,21] for overviews and other examples. Two important papers for our analysis are [13,14]. They provide necessary and sufficient conditions for existence of the steady-state virtual offered waiting time in $G/G/s+GI$ systems. In addition, [13] obtain a closed-form characterization of the (steady-state) distribution of the virtual offered waiting time for the $M/M/s+GI$ system. Because most queueing models with abandonments are not amenable to exact analysis, the heavy traffic asymptotic regime is often used for approximate analysis of such systems. For example, Ward and Glynn [41] study the heavy traffic limit of the $G/G/1+GI$ queue in the conventional heavy traffic limit regime. Ata et al. [12] study a multiclass queue with abandonments under nonlinear costs. Other examples include [11,18,31,34,35]. We refer readers to [40] for a comprehensive survey of the approximate analysis of queueing systems with abandonments.

In contrast, this paper assumes that customers make rational abandonment decisions. Starting with [33], a large and growing literature studying rational customers in queueing systems emerged. The decisions of arriving customers can include whether to join the queue or balk (joining decision) and which queue to join (routing decision). After joining the queue, customers may decide whether to keep waiting in the queue or leave (abandoning or reneging decision) and whether to switch to a different queue (jockeying decision). These models use the equilibrium approach to study the queueing system when customers make rational decisions on their own by imposing consistency conditions on customers' rational decisions. Hassin and Haviv [25] provide a comprehensive survey of this literature.

Customers' abandonment decisions and the resulting system characterization depend on whether they can observe the system state or not; see Chapters 2 and 3 of [25] for a discussion of the observable and unobservable queues, respectively. In the observable case, customers join a physical queue and observe the system state, especially the queue length, and other customers' abandonment decisions. There are several papers studying customers' rational abandonment behavior in the observable case. For example, Afèche and Sarhangian [1] study the rational abandonment decisions of customers in an observable two-class priority system, and pricing as a tool to control this behavior. Other studies include [9,23,30]. In the unobservable setting, customers cannot observe the queue length and other customers' abandonment behavior. A number of papers study how information about the system impacts customers' abandonment behavior. For example, Armony et al. [8] study the system equilibrium using a fluid model under delay announcements. Jennings and Pender [27] study a ticket queue in which a newly arriving customer cannot observe the queue length but knows her ticket number. Moreover, the abandonment decisions are unobservable. They compare this ticket queue with the observable queue and prove that they are indistinguishable in heavy traffic. Kuzu et al. [28] examine customers' abandonment behavior in a ticket queue using a dataset of a Turkish bank. Their empirical study reveals that customers update their forecast of the waiting time over time and adjust their abandonment decisions accordingly. Aksin et al. [4] conduct an empirical study of the impact of different delay announcements on system performance. Examples of unobservable queues include virtual queues in call centers and waiting lists of patients for organ transplant; see, for example, [11]. To the best of our knowledge, the first papers in this literature are [24,26,32,37,43]. These papers assume that the customers determine whether to join the queue and their abandonment time upon arrival. Hassin and Haviv [24] consider a queue in which the reward from service reduces to zero after a certain time and study the Nash equilibrium of customers' strategies. Haviv and Ritov [26] study the symmetric Nash equilibrium of customers' abandonment strategies for a queue in which customers have an increasing and convex waiting cost and a fixed reward from service. In [43]'s model, the customers' abandonment time follows a parametric distribution. The parametric distribution depends only on a single statistical measure of the system performance, for instance the average anticipated waiting time. Mandelbaum and Shimkin [32,37] analyze rational models of a Markovian queue with a general abandonment time distribution (an M/M/s+GI queue). In both papers, the authors assume that the customers make a rational decision on when to abandon upon arrival with the waiting cost and the service utility of the callers given. Callers abandon the system if their actual waiting time exceeds their optimal abandonment time.

This paper studies the endogenous abandonment decisions of customers in the unobservable setting and incorporates such endogenous behavior for a $\text{Geo}/\text{Geo}/1 + GI$ queue. In contrast to the static abandonment models in the literature, we adopt the dynamic abandonment model developed in [3]; see also [4]. In this model, customers decide whether or not to abandon dynamically while they wait in the queue to maximize their utilities. Customers' utilities depend on the waiting costs, the reward from the service and their idiosyncratic random shocks. Customers maximize their utilities by solving an optimal stopping problem. Aksin et al. [3] apply this dynamic abandonment model to the dataset of a US bank call center and study the customers' abandonment behavior in this call center. In equilibrium, customers' belief of the waiting time distribution should be consistent with the actual waiting time distribution. Rather than characterizing the system equilibrium analytically, [3] focus on estimating preference parameters, i.e., the waiting costs and rewards of the customers. The main contribution of our paper is to propose a theoretical framework to study the system equilibrium and the corresponding system performance under the aforementioned endogenous abandonment model. We show that there exists a unique equilibrium and we provide an algorithm to compute it and illustrate its performance.

In particular, our framework for studying the system with endogenous abandonments consists of two parts: an abandonment model and a queueing model. The abandonment model assumes that customers make dynamic abandonment decisions to maximize their expected discounted utility by solving an optimal stopping problem. In particular, the abandonment model characterizes customers' abandonment probability (as a function of how long they have waited so far) given their beliefs of the system performance. The queueing model studies the system dynamics and performance given the customers' abandonment behavior. To be specific, the queueing model characterizes the steady-state distribution of the virtual offered waiting time (VOWT) given customers' abandonment behavior. We restrict our attention to a $\text{Geo}/\text{Geo}/1 + GI$ queue. This enables us to obtain a closed-form characterization of the steady-state distribution of the VOWT. In equilibrium, customers' abandonment behavior (as a function of the system behavior) and the implied system behavior must be consistent with each other; see Definition 1 for a formal definition of the equilibrium.

We show that there exists a unique equilibrium. Showing the existence essentially boils down to invoking the Brouwer–Schauder fixed point theorem in the appropriate space as we do in our analysis. Unfortunately, the existing literature does not provide a satisfactory answer to the question of uniqueness, which is important for computational schemes and also for studying system performance in practice. On the one hand, taking a probabilistic approach, the abstract mathematical question studied in this paper can be posed as one of proving existence and uniqueness of the steady-state distribution of a self-interacting Markov chain (under suitable Markovian assumptions); see [19]. On the other hand, taking a purely analytical approach, it can be formulated as finding the eigenvalues and eigenvectors of a nonlinear mapping which involves "nonlinear Perron–Frobenius theory" [29]. To the best of our knowledge, the literature essentially assumes uniqueness by relying on contraction mapping arguments. However, the fixed point equation we derive to characterize the equilibrium is not a contraction mapping in general. Hence, we establish the uniqueness from first principles.

Our proof of uniqueness is by contradiction. To this end, we first characterize the equilibrium by an auxiliary two-dimensional dynamical system. Different equilibria correspond to different solutions to the dynamical system. Next, we derive the recursive equations that characterize the evolution of this dynamical system (as a function of the waiting time). By studying these recursive equations, we show that they cannot have multiple solutions by contradiction.

We also provide a computational scheme and illustrate the potential value of adopting the aforementioned equilibrium view versus the traditional approach of exogenous modeling of abandonment. The equilibrium is approximated by a truncated one, where the abandonment probabilities of customers who have waited longer than a certain threshold are replaced by an upper bound. The truncated equilibrium converges to the actual equilibrium as the truncation period goes to infinity. Thus, we can approximate the actual equilibrium by the truncated one, which, in turn, is calculated by an iterative algorithm.

The rest of this paper is organized as follows: Sect. 2 characterizes the equilibrium of the $\text{Geo}/\text{Geo}/1 + GI$ queue with endogenous abandonments and states the main theorem at the end. Section 3 proves the main theorem establishing the existence and uniqueness of the equilibrium. Section 4 provides a computational scheme to calculate the equilibrium numerically. We then conduct a numerical study to analyze the impact of modeling the abandonment decisions endogenously in Sect. 5. The appendices are organized as follows: The proofs of the results in Sect. 2, the proofs of lemmas, corollaries and the main theorem in Sect. 3, and the proofs of propositions in Sect. 4 are given in Appendix 1. The proof of a technical lemma that facilitates the proof of the uniqueness of the equilibrium is given in Appendix 2. A road map for the proof of the uniqueness of the equilibrium is in Appendix 3.

## 2 Characterization of the equilibrium

This section introduces the framework to study a single-class queueing with endogenous abandonments. Customers' abandonment behavior is endogenous because their abandonment decisions depend on the congestion in the system. Characterizing such a system requires studying two closely related questions:

1. How do customers' abandonment decisions impact system performance?
2. How does the congestion affect customers' abandonment decisions?

We address the first question by studying a queueing model with abandonments. The queueing model takes customers' abandonment decisions as inputs and studies the resulting waiting time distribution in steady state. We then introduce an abandonment model of forward-looking customers who make abandonment decisions dynamically to maximize their utilities given the underlying system dynamics. The abandonment model takes the waiting time distribution (in steady state) as an input and studies the resulting abandonment time distribution. Combining the queueing and abandonment models gives rise to the framework for studying the single-class queue with endogenous abandonments. To be more specific, we employ an equilibrium approach, which imposes consistency conditions. Namely, it requires that the abandonment and waiting time distributions of the two models must be consistent in equilibrium; see Definition 1. At the end of this section, we state the main theorem which establishes the existence and uniqueness of the equilibrium.

### 2.1 The queueing model

The queueing model characterizes the system performance by taking the distribution of customers' abandonment time as given. We consider a Geo/Geo/1+$GI$ queue in which customers' abandonment times follow a general distribution. In particular, we consider a single-class queue with a single server[1]. In each period, a new customer arrives at the system with probability $a$. We assume a non-idling service policy. Customers are served in a first-come, first-served (FCFS) fashion. The service times follow a geometric distribution with probability $b$. We consider only the underloaded case[2], i.e., $a < b$.

[1] The single-server assumption eliminates the possibility of having multiple service completions in one period, simplifying the characterization of the waiting time distribution significantly in our discrete-time model. Under this assumption, the characterization of the hazard rate of the waiting time distribution is a straightforward analogue of that of [13] in their continuous-time model. Their characterization of the waiting time distribution is valid for multiserver queueing systems as well. This observation leads us to the conjecture that our existence and uniqueness results can be extended to the multiserver setting as well.

[2] Although we restrict attention to the underloaded case, i.e., $a < b$, we conjecture that the existence and uniqueness results continue to hold in the general case when the stability condition $a < bG(\infty)$ holds; see Eq. (1) for the definition of $G$. The intuition for this stems from the fact that the proofs in Sect. 3 rely merely on the properties of the abandonment decisions and performance metrics of the queue when the waiting

A crucial feature of our model is that customers make real-time abandonment decisions while waiting; a detailed description of that will be provided in the next section. Until then, we assume the abandonment time follows a general distribution. To be specific, a customer who has waited in the queue for $w$ periods abandons with probability $q(w)$ for $w \geq 1$.

A quantity of primary interest for us is $\beta(w)$, the steady-state probability of entering service in the next period for a customer who has been in the queue for $w$ periods. This is because the (steady-state) probability $\beta$ of entering service is a key input of customers' abandonment decisions described in Sect. 2.2. We first study the virtual offered waiting time (VOWT) process, denoted by $\{V(t) : t \geq 1\}$, where $V(t)$ is the number of periods a customer arriving at time $t$ has to wait if she does not abandon. Then the (steady-state) probability $\beta$ of entering service is the hazard rate of the steady-state distribution of the VOWT.

Baccelli and Hebuterne [13] calculate the VOWT distribution of an $M/M/s + GI$ queue in closed form. The authors observe that the steady-state distributions of the VOWT of the following two systems are equivalent, which is key to their analysis:

- (System 1) A customer calculates her VOWT upon arrival and balk if it exceeds her patience time.
- (System 2) A customer joins the queue regardless and abandons when her patience time expires.

The system of interest for us falls into the second case (System 2). Following this observation, we can study the VOWT of System 2 by analyzing the VOWT of System 1. Baccelli and Hebuterne [13] derive the generalized Lindley's equation [13, Equation 2.1] to characterize the steady-state distribution of the VOWT (of System 1). We adapt their approach to our discrete-time setting. To this end, let $G(w)$ denote the probability that a customer abandons the queue within $w$ periods. Note that

$$G(w) = 1 - \prod_{i=1}^{w}(1 - q(i)), \tag{1}$$

and let $\bar{G}(w) = 1 - G(w)$. Conditioning on the service time of a (potential) customer who arrives in period $t$, we write a recursive equation to characterize the dynamics of the VOWT $V(t)$ (for $t \geq 1$). This equation is the discrete-time analogue of the generalized Lindley's equation in [13], which is given as follows: For $m = 1, 2, \ldots$,

$$V(t+1) = \begin{cases} (V(t) - 1)^+, & \text{w.p. } (1 - a) + aG(V(t)), \\ V(t) - 1 + m, & \text{w.p. } a(1 - G(V(t)))(1 - b)^{m-1}b. \end{cases} \tag{2}$$

Note that the VOWT $V(t)$ is the sum of the service time of all customers (in System 1) at the beginning of period $t$. Thus, if no customers enter the system in period $t$, the

Footnote 2 continued
time is large. Focusing on the underloaded case, i.e., $a < b$, relieves us from the burden of working with different characterizations of the system as a function of the waiting time, and thus simplifies the proof of the main result.

VOWT decreases by one in a non-empty system. If the system is empty, the VOWT remains zero. This happens either when no customer arrives (with probability $1 - a$) or when a customer arrives, finds that the VOWT exceeds her patience time (with probability $aG(V(t))$), and balks. This gives the first case in Eq. 2. The second case represents the scenario in which a customer arrives at the system and does not balk (with probability $a(1-G(V(t)))$) in period $t$. Conditioning on the service time of this arriving customer, we obtain that the VOWT $V(t+1)$ becomes $V(t) - 1 + m$ (with probability $a(1 - G(V(t)))(1 - b)^{m-1}b$).

The system dynamics are fully characterized by the arrival probability $a$, the probability of service completion $b$ and the abandonment probabilities $q(\cdot)$. Therefore, we can characterize how the system evolves in steady state. By analyzing Eq. (2), we obtain Proposition 1, that is, the discrete-time analogue of [13]'s result. It characterizes the (steady-state) probability of entering service $\beta(\cdot)$ given the abandonment probabilities $q(\cdot)$; see "Proofs of results in Sect. 2" in Appendix 1 for its proof.

**Proposition 1** *The probability of entering service after waiting for $w$ periods in steady state is given as follows:*

$$\beta(w) = \left(1 + \sum_{t=w+1}^{\infty} \prod_{i=w+1}^{t} \frac{1-b}{1 - a\bar{G}(i)}\right)^{-1}, \; w \geq 1. \tag{3}$$

The following recursive equation is immediate from Eq. (3):

$$\frac{1}{\beta(w)} = 1 + \frac{1-b}{1 - a\bar{G}(w+1)} \frac{1}{\beta(w+1)}. \tag{4}$$

The queueing model characterizes the system performance as if the abandonment probabilities $q(\cdot)$ were known. We complete the characterization of the system by introducing a rational abandonment model in the next subsection. The abandonment model takes the system performance (in steady state) as an input and gives the abandonment probabilities $q(\cdot)$ as the output.

### 2.2 An abandonment model with forward-looking customers

This section introduces the abandonment model, which builds on [3]. In our abandonment model, customers make their abandonment decisions based on their beliefs of the probability of entering service, which also characterizes the system performance.

We consider a dynamic abandonment model. In every period, customers waiting in the queue decide to abandon or stay in the queue. Consider a customer who has been waiting in the queue for $w$ periods. Figure 1 shows the sequence of events that she experiences during the next period. We denote her abandonment decision by $d \in \{0, 1\}$; $d = 0$ corresponds to staying in the queue, whereas $d = 1$ corresponds to abandoning. There is a random shock, denoted by $\varepsilon(w, d)$ associated with each action $d$. The random shock captures the changing and unknown factors that impact customers' abandonment decisions in each period. First, she learns the realizations of

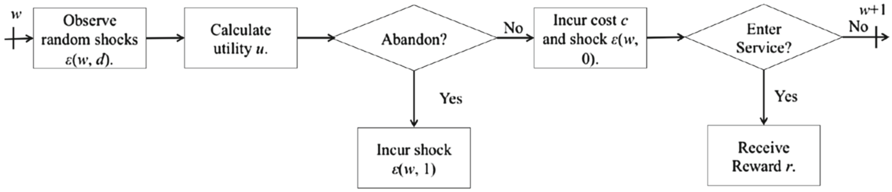


**Fig. 1** Timeline of the events

the random shocks $\varepsilon(w, d)$ and calculates her utility under each decision $d \in \{0, 1\}$, which we denote by $u(w, d, \varepsilon(w, d))$. Then she makes the abandonment decision so as to maximize her utility. If she chooses to abandon ($d = 1$), she only incurs the random shock $\varepsilon(w, 1)$ and leaves the system. If she chooses to stay in the queue ($d = 0$), then she incurs a waiting cost of $c$ for that period as well as the shock $\varepsilon(w, 0)$. Moreover, she enters service at the end of the period with probability $\beta(w)$ in which case she receives a reward $r$ from service. If she does not enter service in that period (which happens with probability $1 - \beta(w)$), she remains in the queue and goes through the same decision making process in the next period. In addition, assume that the customer discounts the utility received in the next period by a factor $\alpha \in [0, 1)$. In making her decision of whether to abandon or wait, she takes into account her expected discounted future utility (or "value-to-go"), denoted by $J(w + 1)$, as well. In summary, the (expected) utility of the customer as a function of her action (at the time of making her abandonment decision) is given as follows:

$$u(w, a, \varepsilon(w, d)) = \begin{cases} -c + \alpha[\beta(w)r + (1 - \beta(w))J(w + 1)] + \varepsilon(w, 0), & \text{if } d = 0, \\ \varepsilon(w, 1), & \text{if } d = 1, \end{cases} \tag{5}$$

where $J(w + 1)$ is the expected discounted future utility of waiting given that she has waited for $w + 1$ periods. Note from Eq. (5) that the utility $u$ merely depends on the waiting cost $c$, reward $r$ and the random shocks $\varepsilon(w, d)$ for $d = 0, 1$. Thus, the random shocks $\varepsilon(\cdot)$ correspond to the other factors that impact customers' abandonment decisions but are not captured by the waiting cost $c$ and the reward $d$.

Equation (5) shows that customers' abandonment decisions solve an optimal stopping problem indeed. Thus, the expected discounted future utility $J(\cdot)$ of waiting is the value function of this optimal stopping problem. To be specific, it is given as follows:

$$J(w) = \mathbb{E}_{\varepsilon(w)} \left[ \max_{d \in \{0,1\}} u(w, a, \varepsilon(w, d)) \right], \tag{6}$$

where $\varepsilon(w) = (\varepsilon(w, 0), \varepsilon(w, 1))$ for $w \geq 1$ and $\mathbb{E}_\varepsilon$ is the expectation over the distribution of $\varepsilon$. The customer's optimal action $d^*(w, \varepsilon(w))$ is given by

$$d^*(w, \varepsilon(w)) = \arg \max_{d \in \{0,1\}} u(w, d, \varepsilon(w, d)). \tag{7}$$

We make the following two assumptions on the idiosyncratic shocks.

**Assumption 1** *The idiosyncratic shocks satisfy the following:*

1. *The idiosyncratic shocks are i.i.d. with zero mean for all $w \geq 1$ and $d \in \{0, 1\}$, i.e., $\mathbb{E}[\varepsilon(w, d)] = 0$;*
2. *The values $-c$ and $r$ are in the interior of the support of $F(\cdot)$, where $F(\cdot)$ is the cumulative distribution function of $\varepsilon(1, 1) - \varepsilon(1, 0)$. In addition, $F(\cdot)$ admits a continuous and positive probability density function $f(\cdot)$ on $[-c, r]$.*

For notational brevity, we suppress the dependence of idiosyncratic shocks on the waiting time and write $\varepsilon(d)$, $d \in \{0, 1\}$. By substituting (5) into (6), we obtain the following Bellman equation:

$$J(w) = \mathbb{E}_\varepsilon \left[\max\{\varepsilon(1), -c + \alpha[\beta(w)r + (1 - \beta(w))J(w + 1)] + \varepsilon(0)\}\right], \ w \geq 1. \tag{8}$$

In light of (7), the probability that a customer abandons the queue after waiting for $w$ periods, $q(w)$, is given as follows:

$$q(w) = \mathbb{P}(d^*(w, \varepsilon) = 1) = \mathbb{P}(u(w, 1, \varepsilon(1)) \geq u(w, 0, \varepsilon(0))), \ w \geq 1. \tag{9}$$

The following lemma shows that the abandonment probability $q(\cdot)$ is uniquely characterized by the expected discounted utility $J(\cdot)$ and vice versa; see "Proofs of results in Sect. 2" in Appendix 1 for its proof.

**Lemma 1** *Given the expected discounted utility $J(\cdot)$, the abandonment probability $q(\cdot)$ is given as follows:*

$$q(w) = \bar{F}\left(-c + \alpha\left\{\beta(w)r + (1 - \beta(w))J(w + 1)\right\}\right), \ w \geq 1, \tag{10}$$

*where $\bar{F}(\cdot) = 1 - F(\cdot)$. In addition, given the abandonment probability $q(\cdot)$, the expected discounted utility $J(\cdot)$ is characterized by the following equation:*

$$J(w) = \mathbb{E}_\varepsilon \left[\bar{F}^{-1}(q(w)) + \varepsilon(0) - \varepsilon(1)\right]^+, \ w \geq 1. \tag{11}$$

In addition, we make the following assumption on the idiosyncratic shocks, the service reward $r$ and the per-period waiting cost $c$.

**Assumption 2** *Customers prefer receiving service immediately to waiting for one period before entering service, i.e., $r > \mathbb{E}_\varepsilon[\max(\varepsilon(1), -c + \alpha r + \varepsilon(0))]$.*

The following proposition (see "Proofs of results in Sect. 2" in Appendix 1 for its proof) shows that the expected value of waiting $J(\cdot)$, and thus the abandonment probability $q(\cdot)$, is unique given the probability of entering service $\beta(\cdot)$.

**Proposition 2** *Given $\beta(\cdot)$, the expected value of waiting $J(\cdot)$ for a customer is the unique solution to the Bellman Eq. (8). Moreover, the corresponding abandonment probability $q(\cdot)$ is uniquely characterized by Eq. (10).*

Building on Lemma 1, Proposition 2 and Assumptions 1 and 2, the following corollary shows that for any given $\beta$, the abandonment probability is bounded away from zero; see "Proofs of results in Sect. 2" in Appendix 1 for its proof.

**Corollary 1** *For any given $\beta(\cdot)$, we have that $J(w) \leq r$ for all $w \geq 1$. Moreover, $q(w) \geq \underline{q} = \bar{F}(r) > 0$ for all $w \geq 1$.*

### 2.3 Existence and uniqueness of the system equilibrium

In an equilibrium, the probability of entering service $\beta$ and the abandonment probability $q$ must be consistent with each other. To facilitate the formal definition of the equilibrium, let $\Omega = [0, 1]^{\infty}$. Note that the probability of entering service $\beta(w)$ and the abandonment probability $q(w)$ (for $w \geq 1$) can be viewed as infinite-dimensional vectors. Thus, $\beta, q \in \Omega$. We let $\Phi : \Omega \to \Omega$ denote the mapping from the vector $q$ of the abandonment probability to the vector $\beta$ of the probability of entering service, which is characterized by Proposition 1. Equations (8) and (10) provide the characterization of the mapping from the vector $\beta$ of the probability of entering service to the vector $q$ of the abandonment probability. We denote that mapping by $\Gamma : \Omega \to \Omega$.

The following definition of system equilibrium imposes the consistency requirement on customers' abandonment probability $q$ and their probability of entering service $\beta$.

**Definition 1** We say that $e^* = (\beta^*, q^*)$ is a system equilibrium (in steady state) if the following conditions are both satisfied:

1. The customers make abandonment decisions with the (steady-state) probability of entering service $\beta^*$, i.e., $q^* = \Gamma(\beta^*)$.
2. The (steady-state) probability of entering service $\beta^*$ is consistent with the actual probability of entering service with abandonment probability $q^*$, i.e., $\beta^* = \Phi(q^*)$.

The definition of the system equilibrium requires that the customers' beliefs on the system performance, which is characterized by $\beta$, are consistent with the actual system performance in steady state[3]. The following corollary, which combines the analysis from the abandonment and queueing models, provides the characterization of the system in equilibrium. It is immediate from Propositions 1 and 2.

**Corollary 2** *The equilibrium is characterized by Eqs.* (1), (3), (8) *and* (10).

We end this section by stating our main result which establishes the existence and uniqueness of the equilibrium. The next section proves this theorem.

**Theorem 1** *There exists a unique system equilibrium $e^*$.*

## 3 Proof of Theorem 1

The proof of Theorem 1 includes two parts. Section 3.1 establishes the existence of the equilibrium, building on several auxiliary lemmas. Section 3.2 proves the uniqueness.

[3] The definition of the system equilibrium is a symmetric Nash equilibrium with infinitely many indistinguishable players; see 1.1 of Chapter 1 in [25] for a detailed discussion.

### 3.1 The existence of the equilibrium

Let $e^* = (\beta^*, q^*)$ be an equilibrium. By definition of an equilibrium, the probability $\beta^*(\cdot)$ of entering service is the solution to the fixed point problem $\beta^* = \Phi(\Gamma(\beta^*))$. We use Brouwer–Schauder fixed point theorem to prove the existence the equilibrium. We first state two lemmas in preparation for applying the fixed point theorem. We then prove the existence result at the end of this subsection.

The following lemma provides an upper bound and a lower bound for the belief $\beta^*(\cdot)$; see "Proofs of results in Sect. 3" in Appendix 1 for its proof.

**Lemma 2** *Given $\beta(w) \in [0, b]$ for $w \geq 1$, let $\tilde{\beta} = (\tilde{\beta}(w) : w \geq 1)$ be defined as $\tilde{\beta} = \Phi(\Gamma(\beta))$. Then $\tilde{\beta}(w)$ is increasing in $w$ and satisfies the following inequality:*

$$\frac{b - a(1-\underline{q})^{w+1}}{1 - a(1-\underline{q})^{w+1}} \leq \tilde{\beta}(w) \leq b \ \ for \ \ w \geq 1, \tag{12}$$

*where $\underline{q} \in (0, 1)$ is the constant given in Corollary 1.*

Since the probability $\beta^*(\cdot)$ of entering service in equilibrium satisfies $\beta^* = \Phi(\Gamma(\beta^*))$, it also satisfies Eq. (12). To state the next lemma, define the set of infinite sequences $\mathscr{B}$ as follows:

$$\mathscr{B} = \left\{ \beta \in l^\infty : \frac{b - a(1-\underline{q})^{w+1}}{1 - a(1-\underline{q})^{w+1}} \leq \beta(w) \leq b, w \geq 1 \right\}, \tag{13}$$

where $l^\infty$ is the space of infinite sequences endowed with the topology induced by the sup-norm. Lemma 3 shows that the fixed point problem $\beta^* = \Phi(\Gamma(\beta^*))$ satisfies the conditions of the Brouwer–Schauder fixed point theorem; see "Proofs of results in Sect. 3" in Appendix 1 for its proof.

**Lemma 3** *The set $\mathscr{B}$ is compact in $l^\infty$. In addition, the mapping $\Phi(\Gamma(\cdot))$ is continuous.*

Thus, we immediately obtain the existence result stated in the following proposition.

**Proposition 3** *There exists a system equilibrium $e^*$.*

*Proof* By Lemma 2, we can restrict our attention to $\beta \in \mathscr{B}$. By Lemma 3, the set $\mathscr{B}$ is compact in $l^\infty$ and the mapping $\Phi(\Gamma(\cdot))$ is continuous. By the Brouwer–Schauder fixed point theorem [42], the fixed point problem $\beta = \Phi(\Gamma(\beta))$ has a solution $\beta^* \in \mathscr{B}$. Therefore, $(\beta^*, \Gamma(\beta^*))$ is a system equilibrium.

### 3.2 The uniqueness of the equilibrium

The uniqueness is proved by contradiction. We first state some properties of the probability $\beta^*(\cdot)$ of entering service and the abandonment probability $q^*(\cdot)$ in equilibrium. We then assume that there exist multiple equilibria. We explore the properties of the difference of two different equilibria. The properties of the difference contradict the other properties of $\beta^*(\cdot)$ and $q^*(\cdot)$ in equilibrium alluded to immediately above.

Intuitively, as a customer waits in the queue, there remain fewer customers ahead of her both due to service completions and abandonments. Thus, her probability of entering service in the next period increases with waiting time. Also, note that for a customer at the head of the queue, $\beta = b$, which is formalized in the next corollary.

**Corollary 3** *In equilibrium, the probability $\beta^*(w)$ of entering service (after waiting for $w$ periods) is increasing in $w$ and satisfies inequality (12). Moreover, we have that*

$$\lim_{w\to\infty} \beta^*(w) = b.$$

The next lemma shows the monotonicity of the expected value of waiting and abandonment probability in equilibrium.

**Lemma 4** *In equilibrium, the expected discounted utility of waiting $J^*(w)$ is increasing, whereas the abandonment probability $q^*(w)$ is decreasing in the waiting time $w$.*

Since the expected value of waiting $J^*$ is increasing and bounded above by $r$, it converges as $w$ tends to infinity. Hence, we have the following corollary.

**Corollary 4** *There exist constants $J_\infty \le r$ and $q_\infty \ge \underline{q}$ such that*

$$\lim_{w\to\infty} J^*(w) = J_\infty \ \ \textit{and} \ \ \lim_{w\to\infty} q^*(w) = q_\infty,$$

*where $J_\infty$ is the unique fixed point of following equation:*

$$x = \mathbb{E}_\varepsilon[-c + \alpha[br + (1-b)x] - (\varepsilon(1) - \varepsilon(0))]^+, \tag{14}$$

*and $q_\infty = \bar{F}(-c + \alpha[br + (1-b)J_\infty])$.*

To construct the contradiction, we assume that there exist at least two different equilibria. Suppose $e_1^*$ and $e_2^*$ are two different equilibria where $e_i^*(w) = (\beta_i^*(w), q_i^*(w))$ for $w \ge 1$ and $i = 1, 2$. Define

$$\delta_\beta(w) = \beta_1^*(w) - \beta_2^*(w),\ \delta_q(w) = q_1^*(w) - q_2^*(w) \ \text{ and } \ \delta_{\bar{G}}(w) = \bar{G}_1^*(w) - \bar{G}_2^*(w).$$

Recall from (1) that $\bar{G}_i^*(w) = \prod_{k=1}^{w}(1-q_i^*(k))$ for $w \ge 1$ and $i = 1, 2$. It is immediate from Corollaries 3 and 4 that

$$\lim_{w\to\infty} \delta_\beta(w) = 0 \ \text{ and } \ \lim_{w\to\infty} \delta_q(w) = 0. \tag{15}$$

To reach a contradiction, we show that (15) cannot hold. Lemmas 5–8 facilitate this contradiction argument. The following lemma proves an auxiliary relationship between $\delta_{\bar{G}}(w)$, $\delta_\beta(w)$ and $\delta_q(w)$, which plays a key role in proving Lemmas 6–7.

**Lemma 5** *There exist two positive sequences $g_1(w)$ and $g_2(w)$ for $w \geq 1$ with $\lim_{w\to\infty} g_i(w) = 0$ for $i = 1, 2$ such that*

$$\delta_{\bar{G}}(w) = g_1(w)\delta_\beta(w) + g_2(w)\delta_q(w). \tag{16}$$

The proof of Lemma 5 is provided in Appendix 2[4].

The following lemma shows that $e_1^*$ and $e_2^*$ must be everywhere different.

**Lemma 6** *If $e_1^* \neq e_2^*$ are two different equilibria, then $e_1^*(w) \neq e_2^*(w)$ for all $w \geq 1$.*

*Proof* We prove this lemma by contradiction. Suppose this is not true. Thus, there exists $n$ such that $e_1^*(n) = e_2^*(n)$. We first show that $e_1^*(w) = e_2^*(w)$ for $w < n$. We then use the assumption $e_1^*(n) = e_2^*(n)$ to show that $e_1^*(w) = e_2^*(w)$ for $w > n$. Combining these two results, we conclude that this contradicts the assumption that $e_1^* \neq e_2^*$.

It follows from Eq. (16) that $\delta_{\bar{G}}(n) = 0$. In other words, $\bar{G}_1^*(n) = \bar{G}_2^*(n)$. Note that $e_i^*(w)$ and $\bar{G}_i^*(w)$ for $i = 1, 2$ and $w \leq n$ can be computed recursively by Eqs. (4), (10)–(11) and the following equation:

$$\bar{G}_i^*(w-1) = \frac{\bar{G}_i^*(w)}{1 - q_i^*(w)}, \, w \leq n.$$

Thus, we have that

$$e_1^*(w) = e_2^*(w) \text{ and } \bar{G}_1^*(w) = \bar{G}_2^*(w), \; w = 1, \ldots, n.$$

In addition, by inverting Eqs. (1), (4) and (10)–(11), we can calculate $\beta(w)$, $q(w)$ and $\bar{G}(w)$ for $w > n$ by the following equations recursively: For $w \geq n$,

$$q(w+1) = \bar{F}\left[H^{-1}\left(\frac{\bar{F}^{-1}(q(w)) + c - \alpha\beta(w)r}{\alpha(1-\beta(w))}\right)\right], \tag{17}$$

$$\bar{G}(w+1) = \bar{G}(w)(1 - q(w+1)), \tag{18}$$

$$\beta(w+1) = \frac{1-b}{1 - a\bar{G}(w+1)} \frac{\beta(w)}{1-\beta(w)}, \tag{19}$$

where $H(x) = \int_{-\infty}^{x} F(y)\,\mathrm{d}y$ for $x \geq 0$. Thus, we have that $e_1^*(w) = e_2^*(w)$ and $\bar{G}_1^*(w) = \bar{G}_2^*(w)$ for all $w > n$ as well. This gives $e_1^* = e_2^*$, which contradicts the assumption that $e_1^* \neq e_2^*$.

[4] Here is a brief outline of the proof in Appendix 2. To prove this result, we define an auxiliary function $f_w(\cdot)$ in "Definition of the auxiliary function $f_w(\cdot)$" in Appendix 2 implicitly and study its properties (especially the monotonicity and convergence of its partial derivatives as $w$ gets large). This function helps characterize $\bar{G}$ in terms of $\beta$ and $q$. We then apply the mean value theorem to $f_w(\cdot)$ to establish the result in Lemma 5.

To facilitate the analysis to follow, let $\mathscr{M}_n$ denote the set of all $n \times n$ matrices, $x = (x_1, \ldots, x_n)'$ be a $n$-dimensional vector, and $M = [m_{ij}] \in \mathscr{M}_n$. Also let $||\cdot||_\infty$ and $|||\cdot|||_\infty$ denote the vector and matrix norms, respectively, given by

$$||x||_\infty = \max_{1\le i\le n} |x_i| \text{ and } |||M|||_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |m_{ij}|.$$

Building on Lemma 5, the next lemma characterizes the evolution of $(\delta_q(w), \delta_\beta(w))'$ as $w$ evolves.

**Lemma 7** *There exist a matrix $A$ and a sequence of matrices $B(w)$, $w \ge 1$ such that the following hold:*

1. *The sequence of vectors $(\delta_q(w), \delta_\beta(w))'$ for $w \ge 1$ satisfies the following recursive equation:*
$$\begin{bmatrix} \delta_q(w+1) \\ \delta_\beta(w+1) \end{bmatrix} = (A + B(w)) \begin{bmatrix} \delta_q(w) \\ \delta_\beta(w) \end{bmatrix}. \tag{20}$$
2. *The two eigenvalues of the matrix $A$, denoted by $\lambda_1$ and $\lambda_2$, satisfy $\lambda_1 > \lambda_2 > 1$.*
3. $\lim_{w\to\infty} |||B(w)|||_\infty = 0$.

*Proof* To facilitate the proof, define the constants $a_1, a_2, a_3$ as follows:

$$a_1 = \frac{1}{\alpha(1-q_\infty)(1-b)},\ a_2 = \frac{f(\bar{F}^{-1}(q_\infty))(\bar{F}^{-1}(q_\infty) + c - \alpha\beta r)}{\alpha(1-q_\infty)(1-\beta)^2} \text{ and}$$
$$a_3 = \frac{1}{1-b}.$$

In addition, define the matrix $A$ as follows:

$$A = \begin{bmatrix} a_1 & a_2 \\ 0 & a_3 \end{bmatrix}.$$

Let $[\varepsilon_q(w+1), \varepsilon_\beta(w+1)]^T$ denote the difference of two vectors given as follows:

$$\begin{bmatrix} \varepsilon_q(w+1) \\ \varepsilon_\beta(w+1) \end{bmatrix} = \begin{bmatrix} \delta_q(w+1) \\ \delta_\beta(w+1) \end{bmatrix} - A \begin{bmatrix} \delta_q(w) \\ \delta_\beta(w) \end{bmatrix},\ w \ge 1. \tag{21}$$

In addition, define the matrices $B(w)$ as follows: For $w \ge 1$,

$$B(w) = \begin{bmatrix} \frac{\varepsilon_q(w+1)}{|\delta_q(w)| + |\delta_\beta(w)|}\text{sign}(\delta_q(w)) & \frac{\varepsilon_q(w+1)}{|\delta_q(w)| + |\delta_\beta(w)|}\text{sign}(\delta_\beta(w)) \\ \frac{\varepsilon_\beta(w+1)}{|\delta_q(w)| + |\delta_\beta(w)|}\text{sign}(\delta_q(w)) & \frac{\varepsilon_\beta(w+1)}{|\delta_q(w)| + |\delta_\beta(w)|}\text{sign}(\delta_\beta(w)) \end{bmatrix},$$

where sign$(x)$ is the sign of $x$. It is easy to check that the matrices $A$ and $B(w)$ satisfy Eq. (20).

We complete the proof by showing that $\lim_{w\to\infty} |||B(w)|||_\infty = 0$. That is, for any $\epsilon > 0$, there exists $w_0$ such that for all $w \geq w_0$, $|||B(w)|||_\infty \leq \epsilon$. Note that this is equivalent to showing that for any $\epsilon > 0$, there exists $w_0$ such that, for all $w \geq w_0$,

$$|\varepsilon_q(w+1)| \leq \epsilon(|\delta_q(w)| + |\delta_\beta(w)|) \text{ and } |\varepsilon_\beta(w+1)| \leq \epsilon(|\delta_q(w)| + |\delta_\beta(w)|). \quad (22)$$

The rest of the proof shows (22). Recall that $H(x) = \int_{-\infty}^{x} F(y)\,\mathrm{d}y$ for $x \geq 0$. To facilitate the proof, define the functions $\psi_1(\cdot,\cdot)$ and $\psi_2(\cdot,\cdot)$ as follows:

$$\psi_1(\beta, q) = \bar{F}\left[H^{-1}\left(\frac{\bar{F}^{-1}(q) + c - \alpha\beta r}{\alpha(1-\beta)}\right)\right],$$

$$\psi_2(\beta, \bar{G}) = \frac{1-b}{1-a\bar{G}}\frac{\beta}{1-\beta}.$$

The partial derivatives of the functions $\psi_1(\cdot,\cdot)$ and $\psi_2(\cdot,\cdot)$ are given as follows:

$$\frac{\partial\psi_1(\beta,q)}{\partial\beta} = -\frac{f(\bar{F}^{-1}(\psi_1(\beta,q))}{1-\psi_1(\beta,q)}\frac{\bar{F}^{-1}(q)+c-\alpha\beta r}{\alpha(1-\beta)^2},$$

$$\frac{\partial\psi_1(\beta,q)}{\partial q} = \frac{f(\bar{F}^{-1}(\psi_1(\beta,q)))}{f(\bar{F}^{-1}(q))\alpha(1-\beta)(1-\psi_1(q,\beta))},$$

$$\frac{\partial\psi_2(\beta,\bar{G})}{\partial\beta} = \frac{1-b}{1-a\bar{G}}\frac{1}{(1-\beta)^2},$$

$$\frac{\partial\psi_2(\beta,\bar{G})}{\partial\bar{G}} = \frac{(1-b)\beta}{1-\beta}\frac{a}{(1-a\bar{G})^2}.$$

It follows from Eqs. (17) to (19) that for $w \geq 1$ and $i = 1, 2$,

$$q_i^*(w+1) = \psi_1(q_i^*(w), \beta_i^*(w)) \text{ and } \beta_i^*(w+1) = \psi_2(\beta_i^*(w), \bar{G}_i^*(w+1)).$$

By the mean value theorem [7, Theorem 8.4], we have that (for $w \geq 1$)

$$\delta_q(w+1) = \frac{\partial\psi_1(\beta_1(w), q_1(w))}{\partial q}\delta_q(w) + \frac{\partial\psi_1(\beta_1(w), q_1(w))}{\partial\beta}\delta_\beta(w), \quad (23)$$

$$\delta_\beta(w+1) = \frac{\partial\psi_2(\beta_2(w), \bar{G}_2(w+1))}{\partial\bar{G}}\delta_{\bar{G}}(w+1) + \frac{\partial\psi_2(\beta_2(w), \bar{G}_2(w+1))}{\partial\beta}\delta_\beta(w), \quad (24)$$

where

$$(\beta_1(w), q_1(w)) = c_1(w)(\beta_1^*(w), q_1^*(w)) + (1-c_1(w))(\beta_2^*(w), q_2^*(w)),$$

$$(\beta_2(w), \bar{G}_2(w+1)) = c_2(w)(\beta_1^*(w), \bar{G}_1^*(w+1)) + (1-c_2(w))(\beta_2^*(w), \bar{G}_2^*(w+1)),$$

for some $c_1(w), c_2(w) \in (0, 1)$. It follows from Corollaries 3 and 4 that for $i = 1, 2$,

$$\beta_i^*(w) \to b,\ q_i^*(w) \to q_\infty \text{ and } \bar{G}_i^*(w) \to 0 \text{ as } w \to \infty.$$

Since $(\beta_1(w), q_1(w))$ and $(\beta_2(w), \bar{G}_2(w+1))$ are convex combinations of the equilibrium quantities, the following holds:

$$(\beta_1(w), q_1(w)) \to (b, q_\infty) \text{ and } (\beta_2(w), \bar{G}_2(w+1)) \to (b, 0) \text{ as } w \to \infty.$$

Then it follows from the continuity of the partial derivatives of $\psi_1(\cdot)$ and $\psi_2(\cdot)$ that

$$\begin{aligned}
\lim_{w\to\infty} \frac{\partial \psi_1(\beta_1(w), q_1(w))}{\partial q} &= \frac{\partial \psi_1(b, q_\infty)}{\partial q} = a_1, \\
\lim_{w\to\infty} \frac{\partial \psi_1(\beta_1(w), q_1(w))}{\partial \beta} &= \frac{\partial \psi_1(b, q_\infty)}{\partial \beta} = a_2, \\
\lim_{w\to\infty} \frac{\partial \psi_2(\beta_2(w), \bar{G}_2(w+1))}{\partial \bar{G}} &= \frac{\partial \psi_2(b, 0)}{\partial \bar{G}} = ab, \\
\lim_{w\to\infty} \frac{\partial \psi_2(\beta_2(w), \bar{G}_2(w+1))}{\partial \beta} &= \frac{\partial \psi_2(b, 0)}{\partial \beta} = a_3.
\end{aligned}$$

Combining these with (23)–(24), we conclude that for any $\epsilon > 0$, there exists $w_1$ such that for $w \geq w_1$,

$$\left|\delta_q(w+1) - a_1\delta_q(w) - a_2\delta_\beta(w)\right| \leq \epsilon(|\delta_q(w)| + |\delta_\beta(w)|), \tag{25}$$

$$\left|\delta_\beta(w+1) - ab\delta_{\bar{G}}(w+1) - a_3\delta_\beta(w)\right| \leq \epsilon(|\delta_q(w)| + |\delta_\beta(w)|). \tag{26}$$

In particular, combining (25) with (21) yields that

$$|\varepsilon_q(w+1)| = |\delta_q(w+1) - a_1\delta_q(w) - a_2\delta_\beta(w)| \leq \epsilon(|\delta_q(w)| + |\delta_\beta(w)|),$$

which gives the first inequality in (22).

To complete the proof, we now focus on the second inequality in (22). By Lemma 5, there exists $w_2 \geq w_1$ such that

$$|\delta_{\bar{G}}(w)| \leq \frac{\epsilon}{2}\left(|\delta_q(w)| + |\delta_\beta(w)|\right) \text{ for } w \geq w_2. \tag{27}$$

In addition, note that

$$\begin{aligned}
|\delta_q(w+1)| &= \left|\delta_q(w+1) - a_1\delta_q(w) - a_2\delta_\beta(w) + a_1\delta_q(w) + a_2\delta_\beta(w)\right| \\
&\leq \left|\delta_q(w+1) - a_1\delta_q(w) - a_2\delta_\beta(w)\right| + \left|a_1\delta_q(w) + a_2\delta_\beta(w)\right| \\
&\leq \epsilon(|\delta_q(w)| + |\delta_\beta(w)|) + a_1|\delta_q(w)| + a_2|\delta_\beta(w)| \\
&\leq M_1(|\delta_q(w)| + |\delta_\beta(w)|),
\end{aligned} \tag{28}$$

where $M_1 = \max\{a_1, a_2\} + \epsilon$ and the second inequality follows from (25). Since $\bar{G}_1^*(w+1) \to 0$ as $w \to \infty$, there exists $w_3 \geq w_2$ such that

$$\bar{G}_1^*(w+1) \leq \frac{\epsilon}{2M_1} \quad \text{for} \quad w \geq w_3. \tag{29}$$

Combining Eqs. (28)–(29), we have the following inequality:

$$|\bar{G}_1^*(w+1)\delta_q(w+1)| \leq \frac{\epsilon}{2}(|\delta_q(w)| + |\delta_\beta(w)|) \quad \text{for} \quad w \geq w_3. \tag{30}$$

Thus, by substituting (18) into the definition of $\delta_{\bar{G}}$, we have that for $w \geq w_3$,

$$\begin{aligned}
|\delta_{\bar{G}}(w+1)| &= |\bar{G}_1^*(w+1) - \bar{G}_2^*(w+1)| \\
&= |\bar{G}_1^*(w)(1 - q_1^*(w+1)) - \bar{G}_2^*(w)(1 - q_2^*(w+1))| \\
&= |\bar{G}_1^*(w)(1 - q_1^*(w+1)) - \bar{G}_1^*(w)(1 - q_2^*(w+1)) \\
&\quad + \bar{G}_1^*(w)(1 - q_2^*(w+1)) - \bar{G}_2^*(w)(1 - q_2^*(w+1))| \\
&= | - \bar{G}_1^*(w)\delta_q(w+1) + (1 - q_2^*(w+1))\delta_{\bar{G}}(w)| \\
&\leq |\bar{G}_1^*(w)\delta_q(w+1)| + |(1 - q_2^*(w+1))\delta_{\bar{G}}(w)| \\
&\leq |\bar{G}_1^*(w)\delta_q(w+1)| + |\delta_{\bar{G}}(w)| \\
&\leq \frac{\epsilon}{2}(|\delta_q(w)| + |\delta_\beta(w)|) + \frac{\epsilon}{2}(|\delta_q(w)| + |\delta_\beta(w)|) \\
&= \epsilon(|\delta_q(w)| + |\delta_\beta(w)|).
\end{aligned} \tag{31}$$

The first inequality simply follows from $|-a+b| \leq |a| + |b|$ for any values $a, b$. The second inequality holds because $|1 - q_2^*(w+1)| \leq 1$. The last inequality follows from Eqs. (27) to (30). In summary, we can rewrite (31) as follows: For $w \geq w_3$,

$$|\delta_{\bar{G}}(w+1)| \leq \epsilon(|\delta_q(w)| + |\delta_\beta(w)|).$$

Then we observe the following for $w \geq w_3$,

$$\begin{aligned}
\left|\delta_\beta(w+1) - a_3\delta_\beta(w)\right| &= \left|\delta_\beta(w+1) - a_3\delta_\beta(w) - ab\delta_{\bar{G}}(w+1) + ab\delta_{\bar{G}}(w+1)\right| \\
&\leq \epsilon(|\delta_q(w)| + |\delta_\beta(w)|) + ab|\delta_{\bar{G}}(w+1)| \\
&\leq 2\epsilon(|\delta_q(w)| + |\delta_\beta(w)|),
\end{aligned} \tag{32}$$

where the first inequality follows from (26), while the last inequality follows because $ab \leq 1$. Thus, it follows from Eqs. (21) to (32) that

$$|\varepsilon_\beta(w+1)| = |\delta_\beta(w+1) - a_3\delta_\beta(w)| \leq 2\epsilon(|\delta_q(w)| + |\delta_\beta(w)|),$$

which gives the second inequality in (22).

The following technical lemma facilitates the proof of uniqueness.

**Lemma 8** *Let $x(w)$ for $w \geq 1$ be a sequence of vectors in $\mathbb{R}^n$ with $x(w) \neq 0$ for all $w \geq 1$ such that*

$$x(w+1) = (A + B(w))x(w), \tag{33}$$

*where $A \in \mathscr{M}_n$ with eigenvalues $\lambda_1 > \cdots > \lambda_n > 1$ and $B(w) \in \mathscr{M}_n$ with* $\lim_{w\to\infty} |||B(w)|||_\infty = 0$*. Then $x(w)$ cannot converge to zero, i.e., $x(w) \nrightarrow 0$ as $w \to \infty$.*

*Proof* We can write (33) equivalently as

$$x(w) = (A + B(w))^{-1}x(w+1)$$

for $w$ large enough because $A + B(w)$ is invertible for large $w$. This is because the eigenvalues of $A$ are larger than one, and $B(w)$ is negligible for large $w$. Defining

$$C(w) = (A + B(w))^{-1} - A^{-1}, \tag{34}$$

we first show that

$$|||C(w)|||_\infty \to 0 \text{ as } w \to \infty.$$

Note that for $w \geq 1$,

$$\begin{aligned} C(w) &= - A^{-1} + (A + B(w))^{-1} \\ &= - A^{-1}\left[I - A(A + B(w))^{-1}\right] \\ &= - A^{-1}\left[(A + B(w))(A + B(w))^{-1} - A(A + B(w))^{-1}\right] \\ &= - A^{-1}(A + B(w) - A)(A + B(w))^{-1} \\ &= - A^{-1}B(w)(A + B(w))^{-1}. \end{aligned} \tag{35}$$

Therefore, the following holds: For $w \geq 1$,

$$\begin{aligned} |||C(w)|||_\infty &= |||A^{-1}B(w)(A + B(w))^{-1}|||_\infty \\ &\leq |||A^{-1}|||_\infty |||B(w)|||_\infty |||(A + B(w))^{-1}|||_\infty. \end{aligned} \tag{36}$$

It follows from (35) that $(A+B(w))^{-1} = A^{-1} - A^{-1}B(w)(A+B(w))^{-1}$. Therefore, we have that (for $w \geq 1$)

$$\begin{aligned} |||(A + B(w))^{-1}|||_\infty &\leq |||A^{-1}|||_\infty + |||A^{-1}B(w)(A + B(w))^{-1}|||_\infty \\ &\leq ||A^{-1}|||_\infty + |||A^{-1}|||_\infty |||B(w)|||_\infty |||(A + B(w))^{-1}|||_\infty. \end{aligned}$$

By rearranging the terms, we obtain that (for $w \geq 1$)

$$|||(A + B(w))^{-1}|||_\infty \leq \frac{|||A^{-1}|||_\infty}{1 - |||A^{-1}|||_\infty |||B(w)|||_\infty}.$$

Substituting this inequality into (36), we have that (for $w \geq 1$)

$$|||C(w)|||_\infty \leq |||A^{-1}|||_\infty \frac{|||A^{-1}|||_\infty}{1 - |||A^{-1}|||_\infty |||B(w)|||_\infty} |||B(w)|||_\infty.$$

Since $|||B(w)|||_\infty \to 0$, we conclude that $|||C(w)|||_\infty \to 0$ as $w \to \infty$.

Next, we show that $x(w) \not\to 0$ by contradiction. Suppose that $x(w) \to 0$. Note that the eigenvalues of $A^{-1}$ are $\lambda_1^{-1} < \cdots < \lambda_n^{-1} < 1$. Thus, $A^{-1}$ is diagonalizable and there exists a matrix $S$ (of the eigenvectors) such that $A^{-1} = S^{-1} \Lambda S$, where $\Lambda$ is a diagonal matrix with diagonal entries $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$. Or, equivalently, we can write that $\Lambda = SA^{-1}S^{-1}$. Define $||\cdot||_S$ and $|||\cdot|||_S$ as follows:

$$||x||_S = ||Sx||_\infty \text{ and } |||M|||_S = |||SMS^{-1}|||_\infty.$$

Therefore, the following holds: For $w \geq 1$,

$$\begin{aligned}
|||A^{-1} + C(w)|||_S =& |||S(A^{-1} + C(w))S^{-1}|||_\infty \\
=& |||\Lambda + SC(w)S^{-1}|||_\infty \\
\leq& |||\Lambda|||_\infty + |||SC(w)S^{-1}|||_\infty \\
\leq& \lambda_n^{-1} + |||S|||_\infty |||S^{-1}|||_\infty |||C(w)|||_\infty,
\end{aligned}$$

where the last inequality follows from $|||\Lambda|||_\infty = \lambda_n^{-1} < 1$. Since $|||C(w)|||_\infty \to 0$, the second term on the right-hand side tends to zero as $w \to \infty$. Thus, there exists $w_0$ and $\epsilon > 0$ such that for $w \geq w_0$,

$$||||A^{-1} + C(w)|||_S \leq \lambda_n^{-1} + \epsilon < 1.$$

Therefore, it follows from (34) that for $w \geq w_0$,

$$\begin{aligned}
||x(w)||_S =& ||(A + B(w))^{-1} x(w+1)||_S \\
=& ||(A^{-1} + C(w)) x(w+1)||_S \\
=& ||S(A^{-1} + C(w))S^{-1} Sx(w+1)||_\infty \\
\leq& |||S(A^{-1} + C(w))S^{-1}|||_\infty ||Sx(w+1)||_\infty \\
=& |||A^{-1} + C(w)|||_S ||x(w+1)||_S \\
\leq& (\lambda_n^{-1} + \epsilon) ||x(w+1)||_S.
\end{aligned}$$

We proceed by contradiction. Suppose $x(w) \to 0$ as $w \to \infty$. Denote $d = ||x(w_0)||_S$. Since $x(w_0) \neq 0$ by assumption, it holds that $d > 0$. Moreover, since $x(w_0 + n) \to 0$ as $n \to \infty$, there exists $n_0$ such that $||x(w_0 + n)||_S < d$ for all $n \geq n_0$. Therefore, the following holds:

$$d = ||x(w_0)||_S \leq (\lambda_n^{-1} + \epsilon)^n ||x(w_0 + n)||_S \leq (\lambda_n^{-1} + \epsilon)^n d < d.$$

This contradiction shows that $x(w)$ cannot converge to zero as $w \to \infty$.

It is immediate from Lemma 7 that our problem satisfies the conditions of Lemma 8 (for $n = 2$). This observation facilitates the uniqueness proof; see, the proof of Proposition 4. Next we state the uniqueness result whose proof follows from Corollaries 3 and 4 and Lemmas 5–8.

**Proposition 4** *The system equilibrium $e^*$ is unique.*

*Proof* We prove the uniqueness by contradiction. Suppose $e_1^*$ and $e_2^*$ are two different equilibria. On the one hand, by Corollaries 3 and 4, Eq. (15) holds. On the other hand, the two eigenvalues of matrix $A$ in Lemma 7 are $(1-b)^{-1}$ and $[\alpha(1-q_\infty)(1-b)]^{-1}$, which are different and strictly greater than 1. Thus, it follows from Lemmas 6 and 7 that the sequence $(\delta_q(w), \delta_\beta(w))$ for $w \geq 1$ satisfies the two conditions in Lemma 8. By Lemma 8, either $\lim_{w\to\infty} \delta_\beta(w) \neq 0$ or $\lim_{w\to\infty} \delta_q(w) \neq 0$, which contradicts Eq. (15).

## 4 An algorithm to compute the equilibrium numerically

This section provides an algorithm to compute the equilibrium. To calculate the equilibrium numerically, we introduce the notion of a truncated equilibrium in which the abandonment decisions are only partially endogenous. The abandonment probability of customers who have waited for more than $N$ periods is given exogenously. They make abandonment decisions as if they had waited in the system indefinitely. Customers who have waited for less than $N$ periods make their abandonment decisions endogenously. Formally, the truncated equilibrium is defined as follows:

**Definition 2** For $N \geq 1$, we call $e_N = (\beta_N, q_N)$ a truncated equilibrium if it satisfies the following conditions:

1. $\beta_N(w) = b$ and $q_N(w) = q_\infty$ for all $w \geq N$.
2. $\beta_N(w) = \Phi(q_N)(w)$ and $q_N(w) = \Gamma(\beta_N)(w)$ for all $w < N$,

where $\Phi(\cdot)$ and $\Gamma(\cdot)$ are the mappings given in Definition 1.

Given $N$, the truncated equilibrium $e_N$ is fully characterized if the values of $\beta_N(N)$, $q_N(N)$ and $\bar{G}_N(N)$ are known, where $\bar{G}_N(\cdot) = 1 - G_N(\cdot)$ and $G_N(\cdot)$ is the cdf induced by the abandonment probability $q_N(\cdot)$; see Eq. (1). To be specific, recall from Eqs. (4) and (10)–(11) that the probability of entering service $\beta_N(w)$ and the abandonment probability $q_N(w)$, for all $w = 1, \ldots, N-1$, can be characterized by the following equations recursively:

$$\beta_N(w) = \left(1 + \frac{1-b}{1-a\bar{G}_N(w+1)} \frac{1}{\beta_N(w+1)}\right)^{-1}, \tag{37}$$

$$q_N(w) = \bar{F}\left(-c + \alpha\left\{\beta_N(w)r + (1-\beta_N(w))J_N(w+1)\right\}\right), \tag{38}$$

where

$$J_N(w+1) = \mathbb{E}_\varepsilon\left[\bar{F}^{-1}(q_N(w+1)) + \varepsilon(0) - \varepsilon(1)\right]^+, \tag{39}$$

and

$$\bar{G}_N(w) = \frac{\bar{G}_N(w+1)}{1 - q_N(w+1)}. \tag{40}$$

By the definition of the truncated equilibrium, the values of $\beta_N(N)$ and $q_N(N)$ are given exogenously. Thus, characterizing the truncated equilibrium is equivalent to determining the value of $\bar{G}_N(N)$. The following lemma shows that the truncated equilibrium is unique; see "Proofs of the proposition and the lemma in Section 4" in Appendix 1 for its proof.

**Lemma 9** *There exists a unique truncated equilibrium $e_N$ for $N \geq 1$.*

Corollaries 3 and 4 suggest that the exogenous abandonments in the truncated equilibrium approximate the endogenous abandonment decisions in the (untruncated) equilibrium well for large $N$. The following proposition verifies this intuition and shows that the equilibrium can be approximated by the truncated one closely; see "Proofs of the proposition and the lemma in Section 4" in Appendix 1 for its proof.

**Proposition 5** *The truncated equilibrium $e_N$ converges to the equilibrium $e^*$ uniformly as $N \to \infty$.*

Thus, we use the truncated equilibrium to approximate the equilibrium $e^*$. Fixing the truncation period $N$, we next provide an algorithm to compute the truncated equilibrium $e_N$. As mentioned earlier, the term $\bar{G}_N(N)$ determines the truncated equilibrium through Eqs. (37)–(40). Also note that we must have $\bar{G}_N(0) = 1$ by definition. The idea behind the algorithm is to start with a guess of $\bar{G}_N(N)$ and to recursively calculate $q_N(w)$, $\beta_N(w)$ and $\bar{G}_N(w)$ for $w < N$. If the guess of $\bar{G}(N)$ is correct, then the $\bar{G}(0)$ value calculated recursively must equal 1. Lemma 10 shows that $\bar{G}_N(0)$ is a monotone function of $\bar{G}_N(N)$ (see "Proofs of the proposition and the lemma in Section 4" in Appendix 1 for its proof), and this observation leads to a simple algorithm.

**Lemma 10** *If $\bar{G}^1_N(N) > \bar{G}^2_N(N)$, then $\bar{G}^1_N(0) > \bar{G}^2_N(0)$, where $\bar{G}^1_N(0)$ and $\bar{G}^2_N(0)$ are the values obtained from Eqs. (37) to (40) recursively by substituting $\bar{G}_N(N) = \bar{G}^1_N(N)$ and $\bar{G}_N(N) = \bar{G}^2_N(N)$, respectively.*

By Lemma 10, if $\bar{G}_N(0) < 1$, the true value of $\bar{G}_N(N)$ is greater than the guessed value. So the initial guess must be increased. Otherwise, i.e., $\bar{G}_N(0) > 1$, we should lower the initial guess. This observation is key to the algorithm provided in Table 1.

## 5 A numerical example

This section presents a numerical example to illustrate the effectiveness of the algorithm proposed in Sect. 4. We first compare the result from the numerical computation with the output of a simulation. In addition, we show the importance of modeling abandonments endogenously by comparing the predictions of the model with endogenous abandonments to those of the model with exogenous abandonments. We end this section by studying how parameter changes impact the predictions of the system performance.

**Table 1** The algorithm for calculating the truncated equilibrium

Algorithm 1: The truncated equilibrium in the single-class case.

Initialize: $\bar{G}_N(N) \leftarrow g^0 \in (0,1)$ and $\bar{g} \leftarrow 1$ and $\underline{g} \leftarrow 0$.
Update the value of $\bar{G}_N(N)$:
while $\bar{g} - \underline{g} > \varepsilon$
  Calculate $\beta_N(\cdot)$, $q_N(\cdot)$ and $\bar{G}_N(\cdot)$ via equations (37)-(40).
  if $\bar{G}_N(0) = 1$
    stop
  else
    if $\bar{G}_N(0) > 1$
      $\bar{g} \leftarrow \bar{G}_N(N)$
    else
      $\underline{g} \leftarrow \bar{G}_N(N)$
    end if
  end if
  Pick $g \in (\underline{g}, \bar{g})$ and $\bar{G}_N(N) \leftarrow g$
end while

### 5.1 The setup of the numerical example

Consider the Geo/Geo/1 queue in which customers make abandonment decisions to maximize their utility. The probability of arrival $a$ equals 0.5. In addition, the service rate $b$ equals 0.8. The per-period waiting cost of customers is $c = 2$, and the reward from service is $r = 6$. The idiosyncratic shocks $\varepsilon(0)$ and $\varepsilon(1)$ both follow the type I extreme value distribution, whose cumulative distribution function is given as follows[5]:

$$F_{\varepsilon(0)}(x) = F_{\varepsilon(1)}(x) = e^{e^{(-x)}}, \; x \geq 0.$$

We refer to this setting as the original system.

Suppose that the system then undergoes a change and the service rate is reduced to $b = 0.51$. To predict the new system performance, and, in particular, the abandonment behavior for the new system, we approximate the equilibrium by the truncated one using the truncation period $N = 30$. In addition, in the model with exogenous abandonments, we assume that the distribution of the abandonment times remains the same as the one in the earlier system (with $b = 0.8$).

To compare the two models, we first simulate the system equilibrium corresponding to $b = 0.51$ iteratively. In the simulation, we start by using the equilibrium probability of entering service $\beta_N(\cdot)$ and the probability of abandoning $q_N(\cdot)$ computed via the algorithm given in Table 1. The simulation gives an empirical distribution of the VOWT. We use this empirical distribution as input and update the abandonment time distribution using the model of Sect. 2.2. We then simulate the system again with the updated distributions of the abandonment time and keep updating the distribution of the abandonment times and the VOWT until the simulation converges numerically.

5 This distributional assumption is commonly made in models studying discrete consumer choice, cf. [6]. [3] also make this assumption.

**Table 2** The mean of the VOWT and abandonment time and the fractions of customers that abandon under the simulation, the equilibrium computation and the exogenous model ($a = 0.5$, $b = 0.51$, $c = 2$, $r = 6$)

| | Simulation | Eq. computation | Error (%) | Exogenous model | Error (%) |
|---|---|---|---|---|---|
| VOWT | 2.537 (0.055) | 2.559 | 0.87 | 4.232 | 66.8 |
| Mean abandonment time | 5.68 (0.063) | 5.79 | 1.9 | 13.79 | 142.7 |
| Percentage abandoning | 42.50% (1.6%) | 42.59% | 0.2 | 18.18% | 57.22 |

The numbers in the parentheses are the standard deviation of the statistics

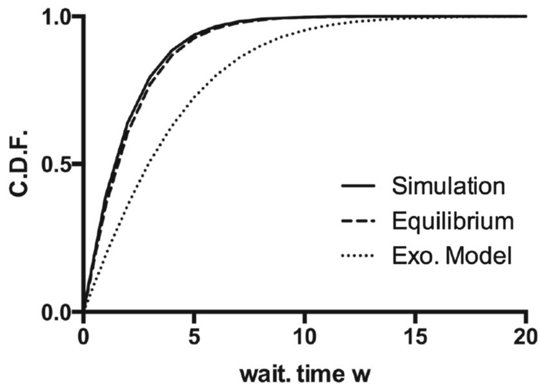


**Fig. 2** The cumulative distribution function of the VOWT with new service rate computed via the simulation, the equilibrium computation and the exogenous model ($a = 0.5$, $b = 0.51$, $c = 2$, $r = 6$)

Table 2 shows the comparison of the means of the VOWT and abandonment time as well as the fraction of customers that abandon in equilibrium calculated from the simulation, the numerical computation and the exogenous model.

In addition, Fig. 2 shows the cumulative distribution function of the VOWT obtained from the simulation, the equilibrium computation and the exogenous model.

Both Table 2 and Fig. 2 show that the exogenous model mistakenly predicts a longer waiting time and a lower probability of abandoning the queue. This is mainly because the model with exogenous abandonments ignores the impact of the service rate change on the abandonment behavior. Under the original service rate (which is higher), the customers are more patient because the probability of entering service is higher. When the service rate drops, the customers are more likely to abandon the system. However, the exogenous abandonment model does not capture this change in customers' behavior. In addition, the comparison of the simulation and the equilibrium computation shows that the proposed truncated equilibrium approximates the equilibrium well in all examples we tried. Thus, we only compare the predictions from the numerical computation of the equilibrium and the exogenous model in the rest of this section.

### 5.2 A comparative statics analysis

We end the numerical study by a comparative statics analysis. To be specific, we study the impact of the changes of the arrival rate $a$ and the service rate $b$ on the predictions of the system performance.

*Impact of the arrival rate* We first study the impact of the arrival rate $a$ on the predictions. We keep the service rate $b$ the same as the original system, i.e., $b = 0.8$, and increase the arrival rate $a$ from 0.5 to 0.79 gradually. Figure 3 shows the numerical characterization of the system equilibrium for three different arrival rates. Figure 3a shows that the abandonment probability (as a function of the waiting time) increases as the arrival rate increases. In other words, customers become more impatient when the system becomes more congested. Since the exogenous model assumes that the abandonment probability remains unchanged, it fails to capture the change in customers' abandonment behavior. Figure 3b shows the probability of entering service in systems with different arrival rates. It shows that the probability of entering service $\beta(\cdot)$ decreases as the arrival rate increases, though customers are more likely to abandon. This is because the system becomes more congested when the arrival rate increases. Figure 3b also shows that the exogenous model underestimates the probability of entering service. A more comprehensive comparison between the equilibrium model and the exogenous model is given in Fig. 4.

Figure 4 shows the comparison of the predictions of the system performance from the equilibrium computation and the exogenous model. It shows that the prediction of the average VOWT from the exogenous model is mistakenly higher when the arrival rate is higher; see Fig. 4a. In addition, Fig. 4b shows that the variance of the VOWT predicted using the exogenous model is higher as well. This is because when the arrival rate is higher, the system becomes more congested. Thus, the customers are more likely to abandon in the more congested system; see Fig. 4c–e. The exogenous model ignores the change in customer's abandonment behavior. Therefore, it underestimates the abandonments and thus results in a higher prediction of the VOWT.

*Impact of the service rate* Next, we study the impact of the service rate $b$ on the predictions of the system performance using a similar method. To be specific, we keep the arrival rate $a$ unchanged, i.e., $a = 0.5$, and reduce the service rate $b$ from 0.8 to 0.51 gradually. Figure 5 shows the system equilibrium for different service rates $b$. Figure 6 compares the predictions of the system performance for the equilibrium computation and the exogenous model. This study also shows that the predictions from the exogenous model are less accurate when the system is more congested, i.e., the service rate $b$ is smaller.

*Impact of both arrival and service rates* The last comparative statics analysis compares the predictions from both the equilibrium computation and the exogenous model under different combinations of the arrival rates and the service rates. Figure 7 shows that the difference of the predictions from the two models is most significant when the arrival rate $a = 0.5$ and the service rate $b = 0.51$. This coincides with the scenario when the system is most congested; see Fig. 7a. This study again emphasizes the importance of the abandonment assumption in congested systems.

## 6 Concluding remarks

In this paper, we study a Geo/Geo/1+$GI$ queue in which the abandonments are endogenous. We characterize the equilibrium of the queue and prove existence and uniqueness of the equilibrium. We also propose an algorithm that computes the system

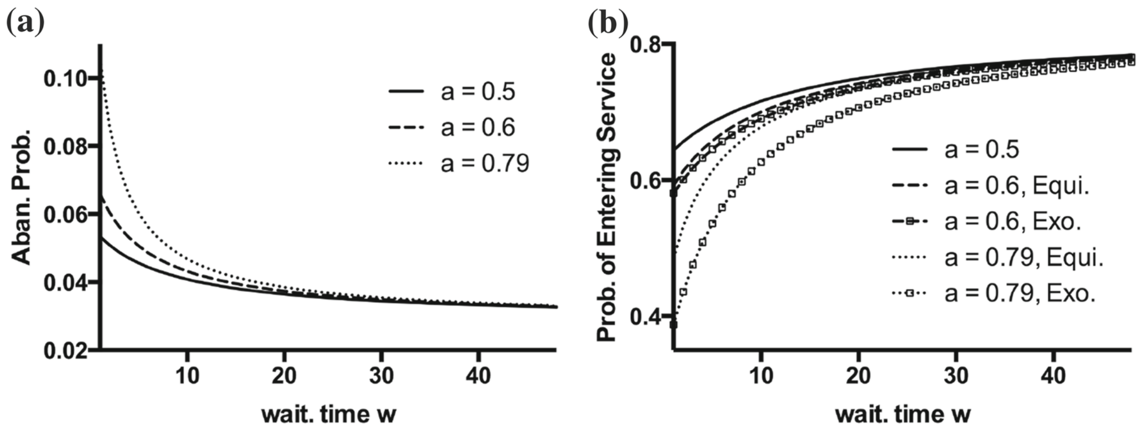


**Fig. 3** The system equilibrium under different arrival rates ($b = 0.8$, $c = 2$ and $r = 6$). **a** The abandonment probability $q(w)$ as a function of the waiting time $w$ . **b** The probability of entering service $\beta(w)$ as a function of the waiting time $w$

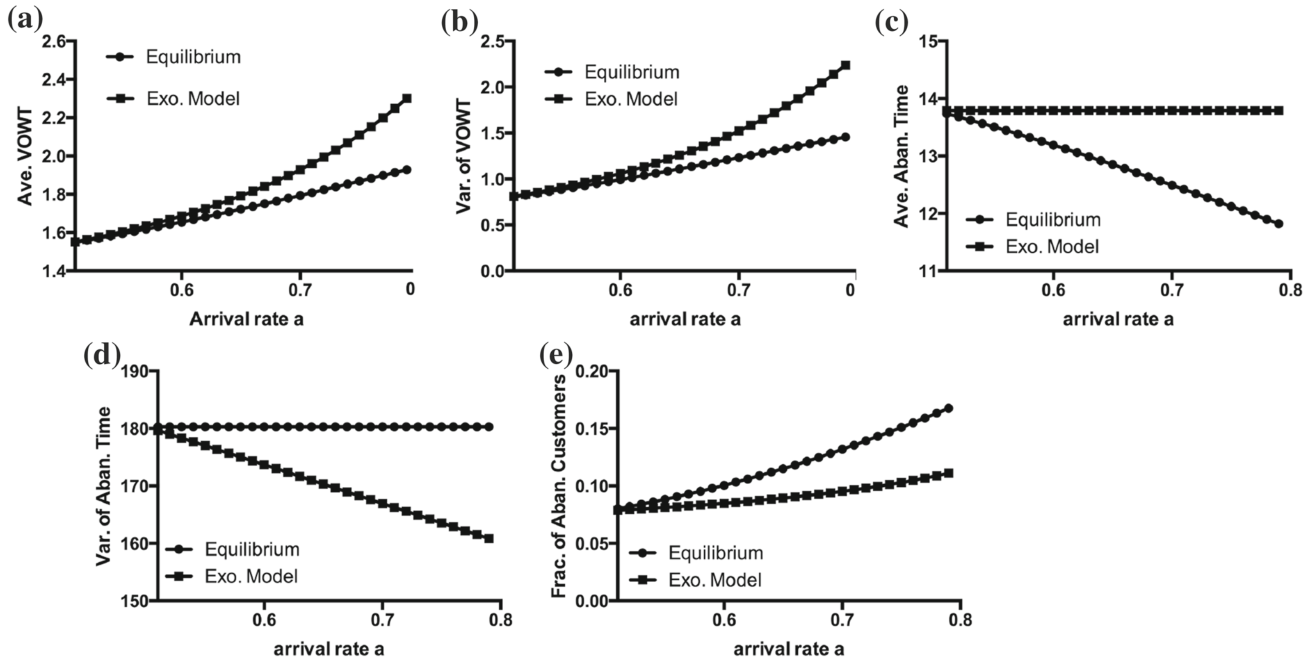


**Fig. 4** The predictions of the system performance under different arrival rates ($b = 0.8$, $c = 2$ and $r = 6$). **a** The average VOWT. **b** The variance of the VOWT. **c** The average abandonment time. **d** The variance of the abandonment time. **e** The fraction of customers abandoning

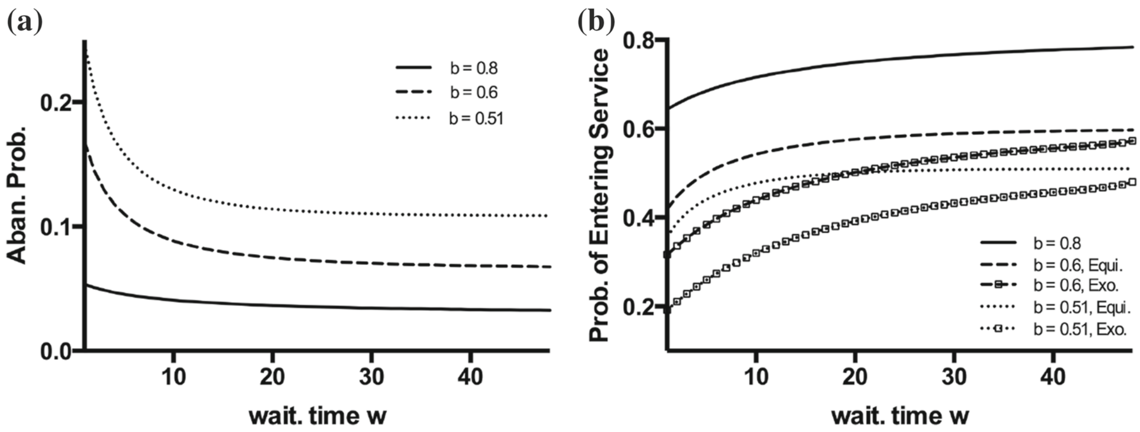


**Fig. 5** The system equilibrium under different service rates ($a = 0.5$, $c = 2$ and $r = 6$). **a** The abandonment probability $q(w)$ as a function of the waiting time $w$. **b** The probability of entering service $\beta(w)$ as a function of the waiting time $w$

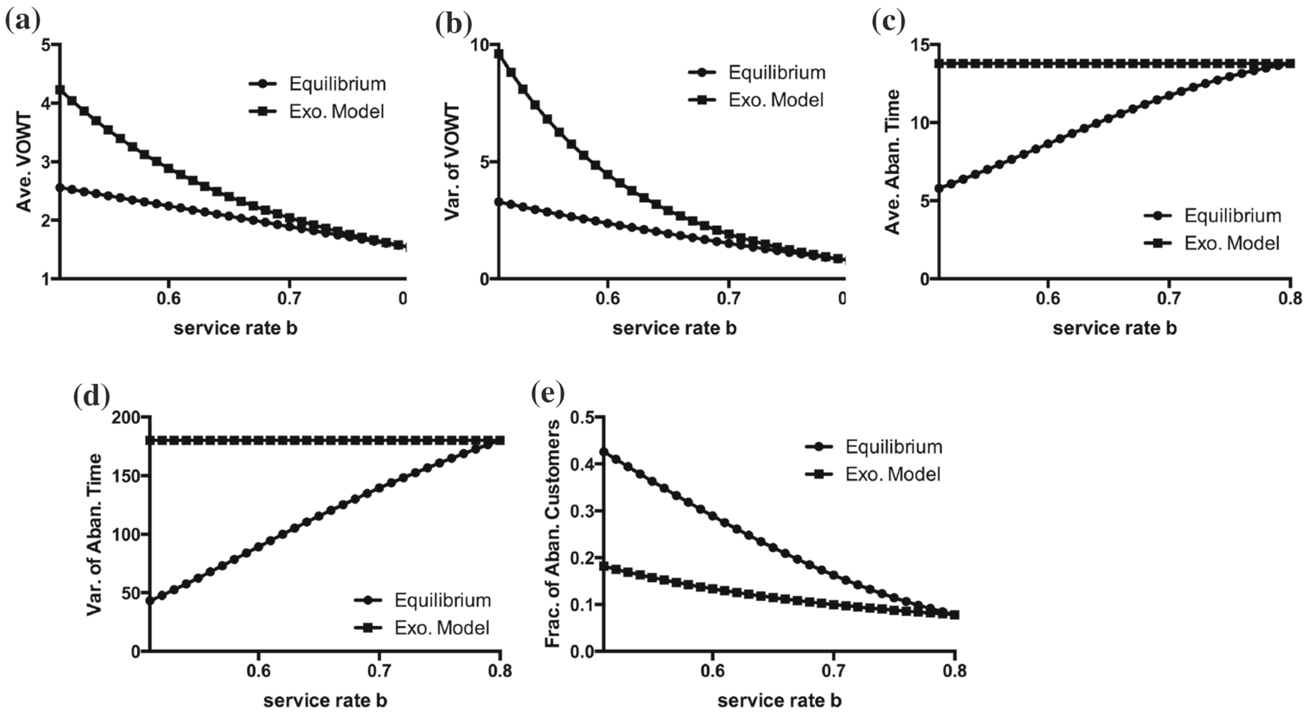


**Fig. 6** The predictions of the system performance under different service rates ($a = 0.5$, $c = 2$ and $r = 6$). **a** The average VOWT. **b** The variance of the VOWT. **c** The average abandonment time. **d** The variance of the abandonment time. **e** The fraction of customers abandoning

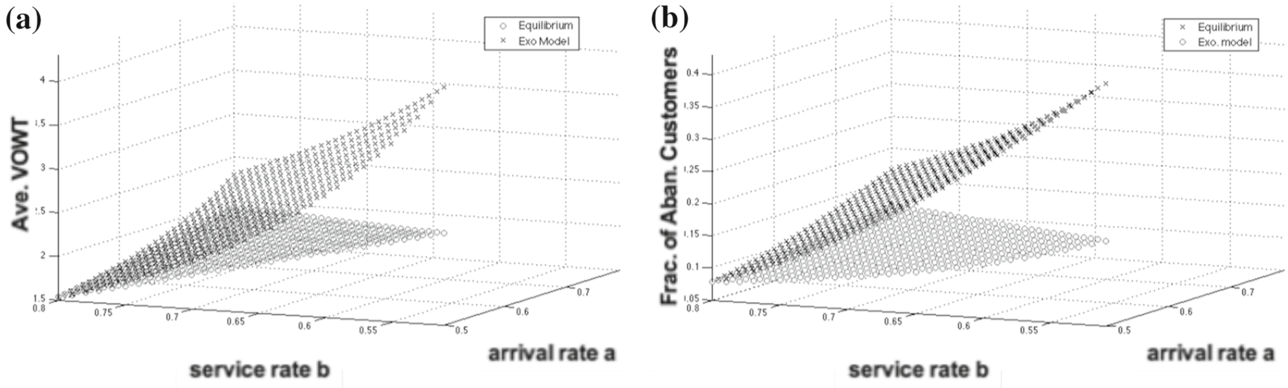


**Fig. 7** The predictions of the system performance under different arrival rates and service rates ($c = 2$ and $r = 6$). **a** The average VOWT. **b** The fraction of customers abandoning

equilibrium. Our analysis points to several future research directions worth exploring. One important question is how to characterize the equilibrium for systems with multiple customer classes. As shown in the numerical system, the impact of the abandonment behavior is more significant when the system is heavily congested. Thus, this problem is more interesting when the system is congested, which motivates the study of the multiple-class system in heavy traffic. This is explored (in the conventional heavy traffic regime) in [10]. Another important direction is to incorporate delay announcements, which is left for future research.

Moreover, although we focus on a single-server queueing system, we conjecture that our existence and uniqueness results will generalize to the multiserver setting. To elaborate on the intuition, note that our single-server assumption eliminates the possibility of having multiple service completions in one period, simplifying the characterization of the waiting time distribution significantly in the discrete-time setting. Under this assumption, the characterization of the hazard rate of the waiting time distribution is a straightforward analogue of that of [13] in their continuous-time model. Their characterization of the hazard rate function of the waiting time distribution is valid for the multiserver queueing systems as well, which motivates our conjecture.

Lastly, note that we assume $a < b$. However, we conjecture that the existence and uniqueness results will generalize under the more general condition that $a < bG(\infty)$. A close look at the proofs of the results in Sect. 3 shows that we merely use the properties of various quantities for large waiting times, where the relevant condition is $a < bG(\infty)$. We assume $a < b$ because it leads to a much simpler analysis, relieving us from the burden of working with different characterizations of the system as a function of the waiting time.

## Appendix 1: Proofs of Lemmas, Propositions, Corollary and Theorems in the paper

### Proofs of results in Sect. 2

*Proof of Proposition 1* It follows from Eq. (2) that $V(t)$ is a Markov process. Since we only consider the underloaded case, i.e., $a < b$, there is a unique stationary distribution of $V(t)$. Let $v(w)$ denote the stationary distribution of $V(t)$, i.e., $v(w) = \lim_{t\to\infty} \mathbb{P}(V(t) = w)$ for $w \geq 1$. The flow balance equations of the Markov process described in (2) are given as follows:

$$v(w) = \begin{cases} \sum_{i=0}^{w} v(i)a\bar{G}(i)(1-b)^{w-i}b + v(w+1)(1 - a\bar{G}(w+1)), & w \geq 1, \\ v(0)(1 - a\bar{G}(0) + ab\bar{G}(0)) + v(1)(1 - a\bar{G}(1)), & w = 0, \end{cases} \tag{41}$$

where $\bar{G}(w) = 1 - G(w)$ for $w \geq 1$. We simplify these equations by defining $r(w) = v(w)(1-b)^{-w}$. By substituting the definition of $v(w)$ into (41) for $w = 0$ and rearranging the terms, we have that

$$(1 - a\bar{G}(1))r(1) - a\bar{G}(0)r(0) = 0. \tag{42}$$

By substituting $r(w)$ into (41), we obtain the following: For $w \geq 1$,

$$\begin{aligned} r(w) =& (1-b)^{-w}\left[\sum_{i=0}^{w}(1-b)^{w}r(i)ab\bar{G}(i)+(1-b)^{w+1}r(w+1)(1-a\bar{G}(w+1))\right] \\ =& \sum_{i=0}^{w} r(i)ab\bar{G}(i) + (1-b)r(w+1)(1-a\bar{G}(w+1)). \end{aligned}$$

Rearranging the terms for $w = 1$, we obtain that

$$(1-b)[(1-a\bar{G}(2))r(2) - r(1)] = b[(1-a\bar{G}(1))r(1) - a\bar{G}(0)r(0)] = 0, \quad (43)$$

where the last equality follows from (42). Subtracting $r(w)$ from $r(w+1)$, we obtain that for $w \geq 1$,

$$\begin{aligned} r(w+1) - r(w) =& (1-b)(1-a\bar{G}(w+2))h(w+2) + r(w+1)ab\bar{G}(w+1) \\ & - (1-b)(1-a\bar{G}(w+1))r(w+1) \\ =& (1-b)(1-a\bar{G}(w+2))r(w+2) \\ & - (1-b-a\bar{G}(w+1))r(w+1). \end{aligned}$$

Rearranging the terms, we have that for $w \geq 1$,

$$\begin{aligned} &(1-b)(1-a\bar{G}(w+2))r(w+2) - r(w+1) \\ &= [(1-a\bar{G}(w+1))r(w+1) - r(w)]. \end{aligned}$$

By substituting (43) into this equation recursively, we have that

$$(1-a\bar{G}(w+1))r(w+1) = r(w), \;\; w \geq 1.$$

This gives that

$$r(w) = r(1)\prod_{i=2}^{w}(1-a\bar{G}(i))^{-1}, w \geq 2.$$

By substituting $v(w) = (1-b)^{w}r(w)$ into this equation, we obtain the following:

$$v(w) = v(1)\prod_{i=2}^{w}\frac{(1-b)}{(1-a\bar{G}(i))}, w \geq 2.$$

Let $\beta(w)$ denote the probability of entering service in next period (in steady state) after waiting for $w$ periods. Thus, $\beta(w)$ is given as follows: For $w \geq 1$,

$$\begin{aligned}\beta(w) &= \lim_{t\to\infty} \mathbb{P}(V(t) = w | V(t) \geq w)\\ &= \frac{v(w)}{\sum_{t=w}^{\infty} v(w)}\\ &= \frac{v(1)\prod_{i=2}^{w}(1-b)(1-a\bar{G}(i))^{-1}}{v(1)\sum_{t=w}^{\infty}\prod_{i=2}^{t}(1-b)(1-a\bar{G}(i))^{-1}}\\ &= \left(1+\sum_{t=w+1}^{\infty}\prod_{i=w+1}^{t}(1-b)(1-a\bar{G}(i))^{-1}\right)^{-1}.\end{aligned}$$

□

*Proof of Lemma 1* Substituting Eq. (5) into (9) yields the following: For $w \geq 1$,

$$\begin{aligned}q(w) &= \mathbb{P}(\varepsilon(1) - \varepsilon(0) \geq -c + \alpha\{\beta(w)r + (1-\beta(w))J(w+1)\})\\ &= \bar{F}(-c + \alpha\{\beta(w)r + (1-\beta(w))J(w+1)\}).\end{aligned}$$

We now derive the equation that characterizes $J(\cdot)$ given $q(\cdot)$. By substituting Eq. (10) into (5), the utility of staying in the queue can be written in terms of the abandonment probability, i.e.,

$$u(w, 0, \varepsilon(w, 0)) = \bar{F}^{-1}(q(w)) + \varepsilon(0).$$

Substituting this into Eq. (6), we obtain that for $w \geq 1$,

$$\begin{aligned}J(w) &= \mathbb{E}_{\varepsilon}\left[\max\{u(w, 0, \varepsilon(w, 0)), u(w, 1, \varepsilon(w, 1))\}\right]\\ &= \mathbb{E}_{\varepsilon}\left[\max\{\bar{F}^{-1}(q(w)) + \varepsilon(0), \varepsilon(1)\}\right]\\ &= \mathbb{E}_{\varepsilon}\left[\bar{F}^{-1}(q(w)) + \varepsilon(0) - \varepsilon(1)\right]^{+} + \mathbb{E}_{\varepsilon}[\varepsilon(1)]\\ &= \mathbb{E}_{\varepsilon}\left[\bar{F}^{-1}(q(w)) + \varepsilon(0) - \varepsilon(1)\right]^{+}.\end{aligned}$$

The last line follows from Assumption 1, i.e., $\mathbb{E}_{\varepsilon}[\varepsilon(1)] = 0$. □

*Proof of Proposition 2* Fixing $\beta(\cdot)$, define $T_{\beta} : l^{\infty} \to l^{\infty}$ as follows: For $w \geq 1$,

$$\begin{aligned}T_{\beta}\, x(w) &= \mathbb{E}_{\varepsilon}[\max\{\varepsilon(1), -c + \alpha[\beta(w)r + (1-\beta(w))x(w+1)] + \varepsilon(0)\}]\\ &= \mathbb{E}_{\varepsilon}[-c + \alpha[\beta(w)r + (1-\beta(w))x(w+1)] - (\varepsilon(1) - \varepsilon(0))]^{+} + \mathbb{E}_{\varepsilon}[\varepsilon(1)]\\ &= \mathbb{E}_{\varepsilon}[-c + \alpha[\beta(w)r + (1-\beta(w))x(w+1)] - (\varepsilon(1) - \varepsilon(0))]^{+},\end{aligned} \tag{44}$$

where $l^\infty$ is the space of bounded sequences of real numbers. The last equality follows from Assumption 1 that $\mathbb{E}_\varepsilon[\varepsilon(1)] = 0$. Note that $T_\beta$ is the right-hand side of Eq. (8). Therefore, the expected discounted utility function $J$ is the fixed point of operator $T_\beta$, i.e., $J = T_\beta J$.

We use Blackwell's sufficient conditions for a contraction [39, Theorem 3.3] to show that the operator $T_\beta$ is a contraction[6].

First, we check that if $J^1(w) \leq J^2(w)$ (for all $w \geq 1$), then $T_\beta J_k^1(w) \leq T_\beta J^2(w)$ for all $w \geq 1$. The following inequality holds: For all $w \geq 1$,

$$(1-\beta(w))J^1_{(w+1)} \leq (1-\beta(w))J^2(w+1),$$

because $1-\beta(w) \geq 0$ and $J^1(w+1) \leq J^2(w+1)$. Since the inequality is preserved by the max operator and the expectation, it follows that $T_\beta J^1(w) \leq T_\beta J^2(w)$ for all $w \geq 1$.

Then we show that $T_\beta (J+e)(w) \leq T_\beta J(w) + \alpha e$ for all $e > 0$ and $w \geq 1$. It follows that

$$\begin{aligned}
&T_\beta (J+e)(w)\\
&\quad = \mathbb{E}_\varepsilon [\max\{\varepsilon(1), -c + \alpha[\beta(w)r + (1-\beta(w))[J(w+1)+e]] + \varepsilon(0)\}]\\
&\quad \leq \mathbb{E}_\varepsilon [\max\{\varepsilon(1), -c + \alpha[\beta(w)r + (1-\beta(w))J(w+1)] + \varepsilon(0) + \alpha e\}]\\
&\quad \leq \mathbb{E}_\varepsilon [\max\{\varepsilon(1) + \alpha e, -c + \alpha[\beta(w)r + (1-\beta(w))J(w+1)] + \varepsilon(0) + \alpha e\}]\\
&\quad = \mathbb{E}_\varepsilon [\max\{\varepsilon(1), -c + \alpha[\beta(w)r + (1-\beta(w))J(w+1)] + \varepsilon(0)\}] + \alpha e\\
&\quad = T_\beta J(w) + \alpha e.
\end{aligned}$$

The first inequality holds because $(1-\beta(w))e \leq e$. Hence, the two sufficient conditions of Blackwell are satisfied, i.e., $T_\beta$ is a contraction mapping. It follows from the Banach fixed point theorem that there exists a unique fixed point of $x = T_\beta x$. Since the solution to Eq. (8) is equivalent to the fixed point of $J = T_\beta J$, the solution is unique. □

*Proof of Corollary 1* Fix $\beta \in [0,1]^\infty$. Let $\mathscr{V} = \{\nu \in l^\infty : \nu(w) \in [0,r],\ w \geq 1\}$. In addition, let $T_\beta : l^\infty \to l^\infty$ be the operator defined in Eq. (44). For any $J^0 \in \mathscr{V}$ and $w \geq 1$, the following inequality holds:

$$\begin{aligned}
T_\beta J^0(w) =&\mathbb{E}_\varepsilon \left[\max\{\varepsilon(1), -c + \alpha[\beta(w)r + (1-\beta(w))J_k^0(w+1)] + \varepsilon(0)\}\right]\\
\leq&\mathbb{E}_\varepsilon [\max\{\varepsilon(1), -c + \alpha r + \varepsilon(0)\}] \leq r.
\end{aligned}$$

The first inequality holds because $J^0 \in \mathscr{V}$. In particular, $J^0(w+1) \leq r$. The second inequality follows from Assumption 2. In addition, it is immediate from (44) that $T_\beta J^0(w) \geq 0$ for all $w \geq 1$. Therefore, $T_\beta J^0 \in \mathscr{V}$. Since $T_\beta$ is a contraction mapping and $\mathscr{V}$ is closed, $J = \lim_{n\to\infty} T_\beta^n J^0 \in \mathscr{V}$. In particular, $J(w) \leq r$ for $w \geq 1$.

[6] Stokey et al. [39] state the result for subsets of $\mathbb{R}^n$ for some integer $n$. The result and proof can be generalized to any subset of a Banach space, especially $(l^\infty, ||\cdot||_\infty)$.

It follows from Eq. (10) to $J(w+1) \leq r$ that

$$q(w) \geq \bar{F}(-c+\alpha r) \geq \bar{F}(r),\ \ w \geq 1.$$

Thus, we have that $q(w) \geq \underline{q} = \bar{F}(r)$. Lastly, we have that $\underline{q} > 0$ because $r$ is in the interior of the support of $\bar{F}(\cdot)$. □

**Proofs of results in Sect. 3**

*Proof of Lemma 2* For a given $\beta \in [0,1]^{\infty}$, let $\tilde{\beta} = \Phi(\Gamma(\beta))$. Note that the mapping $\Phi(\Gamma(\cdot))$ is characterized by Eqs. (8), (10), (1) and (3).

Note from (1) that $\bar{G}(w) \geq 0$ for all $w \geq 1$. Thus, it follows from Eq. (3) that for $w \geq 1$,

$$\tilde{\beta}(w) \leq \left(1+\sum_{i=1}^{\infty}(1-b)^i\right)^{-1} = b.$$

This gives the upper bound of Eq. (12). In addition, it also follows from Eq. (3) that for $w \geq 1$,

$$\begin{aligned}\tilde{\beta}(w) &= \left(1+\sum_{t=w+1}^{\infty}\prod_{i=w+1}^{t}\frac{1-b}{1-a\bar{G}(i+1)}\right)^{-1}\\ &\geq \left(1+\sum_{i=1}^{\infty}\left(\frac{1-b}{1-a\bar{G}(w+1)}\right)^i\right)^{-1}\\ &=1-\frac{1-b}{1-a\bar{G}(w+1)}\\ &=\frac{b-a\bar{G}(w+1)}{1-a\bar{G}(w+1)}.\end{aligned} \tag{45}$$

The inequality follows from the fact that $\bar{G}(w)$ is non-increasing, i.e., $\bar{G}(i) \leq \bar{G}(w+1)$ for all $i \geq w+1$; see (1) for its definition. It follows from Corollary 1 that $q_1(w) \geq \bar{q} > 0$ for all $w \geq 1$. Thus, it follows from Eq. (1) that

$$\bar{G}(w) = \prod_{i=1}^{w}(1-q(i)) \leq (1-\underline{q})^w,\ \ w \geq 1.$$

Substituting this inequality into Eq. (45), we have that for $w \geq 1$,

$$\tilde{\beta}(w) \geq 1-\frac{1-b}{1-a(1-\underline{q})^{w+1}} = \frac{b-a(1-\underline{q})^{w+1}}{1-a(1-\underline{q})^{w+1}}.$$

This shows the lower bound of $\tilde{\beta}(w)$ provided in Eq. (12). We end the proof by showing that $\tilde{\beta}(w)$ is non-decreasing in $w$. Rearranging the terms in Eq. (4), we have the following: For $w \geq 1$,

$$\frac{1}{\tilde{\beta}(w+1)} = \left(\frac{1}{\tilde{\beta}(w)} - 1\right)\left(1 + \frac{b - a\bar{G}(w+1)}{1-b}\right).$$

Substituting this equation into the following, we obtain that for $w \geq 1$,

$$\begin{aligned}\frac{1}{\tilde{\beta}(w+1)} - \frac{1}{\tilde{\beta}(w)} &= \left(\frac{1}{\tilde{\beta}(w)} - 1\right)\left(1 + \frac{b - a\bar{G}(w+1)}{1-b}\right) - \frac{1}{\tilde{\beta}(w)} \\ &= \frac{1}{\tilde{\beta}(w)} \frac{b - a\bar{G}(w+1)}{1-b} - \frac{1 - a\bar{G}(w+1)}{1-b} \\ &\leq \frac{1 - a\bar{G}(w+1)}{b - a\bar{G}(w+1)} \frac{b - a\bar{G}(w+1)}{1-b} - \frac{1 - a\bar{G}(w+1)}{1-b} = 0.\end{aligned}$$

The last inequality follows from (45). Thus, we have that $\tilde{\beta}(w) \leq \tilde{\beta}(w+1)$ for all $w \geq 1$, i.e., $\tilde{\beta}(w)$ is non-decreasing in $w$. □

*Proof of Lemma 3* We first show that $\mathscr{B}$ is a compact set. Define a sequence $x_w$ as follows:

$$x_w = b - \frac{b - a(1-\underline{q})^{w+1}}{1 - a(1-\underline{q})^{w+1}}, \ w \geq 1.$$

Thus, it is equivalent to writing $\mathscr{B}$ as

$$\mathscr{B} = \{\beta \in l^\infty : b - x_w \leq \beta(w) \leq b, w \geq 1\}.$$

Note that $x_w \to 0$ as $w \to \infty$. Thus, for any $\epsilon > 0$, there exists $n$ such that $x_w < \epsilon$ for $w \geq n$. Define a set $\mathscr{B}_n$ as follows:

$$\mathscr{B}_n = \{\beta \in \mathbb{R}^n : b - x_w \leq \beta(w) \leq b, w = 1, \ldots, n\}.$$

Since $\mathscr{B}_n$ is a compact set in $\mathbb{R}^n$, it is totally bounded, i.e., it has a finite cover of open balls of radius $\epsilon$. In other words, there exist $l$ and $\nu_1, \ldots, \nu_l \in \mathbb{R}^n$ such that $\mathscr{B}_n \subseteq \cup_{i=1}^l B_n(\nu_i, \epsilon)$, where $B_n(\nu_i, \epsilon)$ is the open ball in $\mathbb{R}^n$ centered at $\nu_i$ and with radius $\epsilon$. Let $w_i = (\nu_i, 0, \ldots)$, $i = 1, \ldots, l$. It is immediate that $\mathscr{B}$ is covered by $B(w_i, \epsilon)$, $i = 1, \ldots, l$, where $B(w_i, \epsilon)$ is the open ball in $l^\infty$ that centers at $w_i$ and has a radius $\epsilon$. Since $\epsilon$ is arbitrary, $\mathscr{B}$ is totally bounded. Since $l^\infty$ is a complete metric space, the totally bounded subset $\mathscr{B}$ of $l^\infty$ is compact; see Theorem 3.28 in [5].

Next we show that $\Phi(\Gamma(\cdot))$ is continuous. Note that $\Phi(\Gamma(\cdot))$ is characterized by Eqs. (8), (10), (1) and (3). Let $\beta_n, \beta \in \mathscr{B}$ be sequences such that $\beta_n \to \beta$ (under the sup-norm). Let $J_n, q_n, G_n$ and $\tilde{\beta}_n$ and $J, q, G$ and $\tilde{\beta}$ be the left-hand sides of Eqs. (8), (10), (1) and (3) by substituting $\beta_n$ and $\beta$ into $\Phi(\Gamma(\cdot))$, respectively. Thus, we have that $\tilde{\beta}_n = \Phi(\Gamma(\beta_n))$ and $\tilde{\beta} = \Phi(\Gamma(\beta))$. We need to show that $\tilde{\beta}_n \to \tilde{\beta}$ under the sup-norm.

It follows from $\beta_n \to \beta$ that for any $\epsilon > 0$, there exists $n_1$ such that $|\beta_n(w) - \beta(w)| < \epsilon$ for all $n \geq n_1$ and $w \geq 1$. It follows from (8) that for all $w \geq 1$ and $n \geq n_1$,

$$
\begin{aligned}
&|J_n(w) - J(w)| \\
&\quad = |\mathbb{E}_\varepsilon\left[\max\{\varepsilon(1), -c + \alpha[\beta_n(w)r + (1-\beta_n(w))J_n(w+1)] + \varepsilon(0)\}\right] \\
&\qquad - \mathbb{E}_\varepsilon\left[\max\{\varepsilon(1), -c + \alpha[\beta(w)r + (1-\beta(w))J(w+1)] + \varepsilon(0)\}\right]| \\
&\quad = |\mathbb{E}_\varepsilon\left[-c + \alpha[\beta_n(w)r + (1-\beta_n(w))J_n(w+1)] + \varepsilon(0) - \varepsilon(1)\right]^+ \\
&\qquad - \mathbb{E}_\varepsilon\left[-c + \alpha[\beta_n(w)r + (1-\beta(w))J(w+1)] + \varepsilon(0) - \varepsilon(1)\right]^+| \\
&\quad \le |\mathbb{E}_\varepsilon\left[\alpha[\beta_n(w)r + (1-\beta_n(w))J_n(w+1)] - \alpha[\beta(w)r + (1-\beta(w))J(w+1)]\right]| \\
&\quad = \alpha|(r - J_n(w+1))(\beta_n(w) - \beta(w)) - (1-\beta(w))(J_n(w+1) - J(w+1))| \\
&\quad \le \alpha r|\beta_n(w) - \beta(w)| + \alpha|J_n(w+1) - J(w+1)| \\
&\quad \le \alpha r\epsilon + \alpha|J_n(w+1) - J(w+1)|.
\end{aligned}
$$

The equality in the third line follows from the fact that $\mathbb{E}[\varepsilon(1)] = 0$. The inequality in the fourth line follows from $|x_1^+ - x_2^+| \le |x_1 - x_2|$ for all $x_1, x_2 \in \mathbb{R}$. It follows from Corollary 1 that $J_n(w), J(w) \in [0, r]$ for all $w \ge 1$. Thus, $|J_n(w+1) - J(w+1)|$ is bounded for all $w \ge 1$. Thus, by applying this inequality recursively, we obtain that for all $w \ge 1$ and $n \ge n_1$,

$$
|J_n(w) - J(w)| \le \alpha r\epsilon \sum_{i=0}^{\infty} \alpha^i = \frac{\alpha r\epsilon}{1-\alpha}. \tag{46}
$$

Let $C_0 = \sup_{x\in[-c,r]} f(x)$. It follows from Assumption 1 that $f(\cdot)$ is continuous. Thus, $C_0 < \infty$. Since $\bar{F}'(x) = -f(x)$, it holds that for any $x_1, x_2 \in [-c, r]$,

$$
|\bar{F}(x_1) - \bar{F}(x_2)| \le C_0|x_1 - x_2|. \tag{47}
$$

Since $J_n(w) \le r$ and $J(w) \le r$ for all $w \ge 1$, we have that (for $w \ge 1$)

$$
\begin{aligned}
-c &\le -c + \alpha(\beta_n(w)r + (1-\beta_n(w))J_n(w+1)) \le -c + \alpha r \le r, \\
-c &\le -c + \alpha(\beta(w)r + (1-\beta(w))J(w+1)) \le -c + \alpha r \le r.
\end{aligned}
$$

Thus, it follows from (10) to (46) that for all $n \ge n_1$ and $w \ge 1$,

$$
\begin{aligned}
|q_n(w) - q(w)| &\le C_0|\alpha[\beta_n(w)r + (1-\beta_n(w))J_n(w+1)] \\
&\qquad -\alpha[\beta(w) + (1-\beta(w))J(w+1)]| \\
&= C_0\alpha|(r - J_n(w+1))(\beta_n(w) - \beta(w)) \\
&\qquad -(1-\beta(w))(J_n(w+1) - J(w+1))| \\
&\le C_0\alpha r|\beta_n(w) - \beta(w)| + c\alpha|J_n(w+1) - J(w+1)| \\
&\le C_0\alpha r\epsilon + c\alpha\frac{\alpha r\epsilon}{1-\alpha} \\
&= \frac{C_0\alpha r\epsilon}{1-\alpha} = C_1\epsilon,
\end{aligned}
$$

where $C_1 = C_0\alpha r/(1-\alpha)$. The first inequality follows from (47) and the second one follows from $J_n(w+1) \in [0, r]$ and $\beta(w) \in [0, 1]$. The last inequality follows from the assumption that $|\beta_n(w) - \beta(w)| < \epsilon$ for $n \geq n_1$ and (46). Substituting this inequality into (1), we have that for all $n \geq n_1$ and $w \geq 1$,

$$\begin{aligned} |G_n(w+1) - G(w+1)| =& |\bar{G}_n(w+1) - \bar{G}(w+1)| \\ =& |\bar{G}_n(w)(1 - q_n(w+1)) - \bar{G}(w)(1 - q(w+1))| \\ \leq& \bar{G}_n(w)|q_n(w+1) - q(w+1)| \\ &+ (1 - q(w+1))|\bar{G}_n(w) - \bar{G}(w)| \\ \leq& (1 - \underline{q})^w C_1\epsilon + (1 - \underline{q})|\bar{G}_n(w) - \bar{G}(w)|, \end{aligned} \tag{48}$$

where the last inequality follows from Corollary 1 and Eq. (1) that

$$|1 - q(w+1)| \leq 1 - \underline{q},\ 0 \leq \bar{G}_n(w) \leq (1 - \underline{q})^w \text{ and } 0 \leq \bar{G}(w) \leq (1 - \underline{q})^w,\ w \geq 1. \tag{49}$$

Applying (48) recursively, we obtain that for all $w \geq 1$ and $n \geq n_1$,

$$\begin{aligned} |G_n(w) - G(w)| &\leq \sum_{i=1}^{w-1}(1 - \underline{q})^i C_1\epsilon + (1 - \underline{q})^{w-1}|\bar{G}_n(1) - \bar{G}(1)| \\ &\leq \sum_{i=1}^{\infty}(1 - \underline{q})^i C_1\epsilon + (1 - \underline{q})^{w-1}|q_n(1) - q(1)| \\ &\leq \frac{C_1\epsilon}{\underline{q}} + C_1\epsilon = \left(1 + \frac{1}{\underline{q}}\right) C_1\epsilon. \end{aligned}$$

By letting $C_2 = (1 + 1/\underline{q})C_1$, we have that $|G_n(w) - G(w)| \leq C_2\epsilon$ for all $w \geq 1$ and $n \geq n_1$. It follows from (4) that for $w \geq 1$ and $n \geq n_1$,

$$\begin{aligned} &|\tilde{\beta}_n(w) - \tilde{\beta}(w)| \\ &= \tilde{\beta}_n(w)\tilde{\beta}(w)\left|\frac{1}{\tilde{\beta}_n(w)} - \frac{1}{\tilde{\beta}(w)}\right| \\ &\leq b^2\left|\frac{1-b}{(1 - a\bar{G}_n(w+1))\tilde{\beta}_n(w+1)} - \frac{1-b}{(1 - a\bar{G}(w+1))\tilde{\beta}(w+1)}\right| \\ &= b^2(1-b)\frac{|(1 - a\bar{G}_n(w+1))(\tilde{\beta}_n(w+1) - \tilde{\beta}(w+1)) + a\tilde{\beta}(w+1)(\bar{G}_n(w+1) - \bar{G}(w+1))|}{(1 - a\bar{G}_n(w+1))\tilde{\beta}_n(w+1)(1 - a\bar{G}(w+1))\tilde{\beta}(w+1)} \\ &\leq b^2(1-b)\frac{(1 - a\bar{G}_n(w+1))|\tilde{\beta}_n(w+1) - \tilde{\beta}(w+1)| + a\tilde{\beta}(w+1)|\bar{G}_n(w+1) - \bar{G}(w+1)|}{(1 - a\bar{G}_n(w+1))\tilde{\beta}_n(w+1)(1 - a\bar{G}(w+1))\tilde{\beta}(w+1)} \\ &= \frac{b^2(1-b)|\tilde{\beta}_n(w+1) - \tilde{\beta}(w+1)|}{\tilde{\beta}_n(w+1)(1 - a\bar{G}(w+1))\tilde{\beta}(w+1)} + \frac{ab^2(1-b)|\bar{G}_n(w+1) - \bar{G}(w+1)|}{(1 - a\bar{G}_n(w+1))\tilde{\beta}_n(w+1)(1 - a\bar{G}(w+1))}, \end{aligned} \tag{50}$$

where the first inequality follows from Lemma 2 that $\tilde{\beta}_n(w) \le b$, $\tilde{\beta}(w) \le b$ and (4). Note that $(b-a(1-\underline{q})^w)/(1-a(1-\underline{q})^w) \to b$ as $w \to \infty$. In addition, $(1-\underline{q})^w \to 0$ as $w \to \infty$. Thus, there exists $w_1$ such that for $w \ge w_1$,

$$\left(\frac{b-a(1-\underline{q})^w}{1-a(1-\underline{q})^w}\right)^2 (1-a(1-(1-\underline{q})^w)) \ge b^2\sqrt{1-b},$$
$$\frac{b-a(1-\underline{q})^w}{1-a(1-\underline{q})^w}(1-a(1-(1-\underline{q})^w))^2 \ge b\sqrt{1-b}.$$

Substituting these two inequalities into (12) and (49), we have that for $w \ge w_1$ and $n \ge n_1$,

$$\tilde{\beta}_n(w+1)(1-a\bar{G}(w+1))\tilde{\beta}_n(w+1) \ge b^2\sqrt{1-b},$$
$$(1-a\bar{G}_n(w+1))\tilde{\beta}_n(w+1)(1-a\bar{G}(w+1)) \ge b\sqrt{1-b}.$$

Substituting these two inequalities into (50) yields that for $w \ge w_1$ and $n \ge n_1$

$$\begin{aligned}|\tilde{\beta}_n(w)-\tilde{\beta}(w)| \le& \sqrt{1-b}|\beta_n(w+1)-\tilde{\beta}(w+1)| \\ &+ ab\sqrt{1-b}|\bar{G}_n(w+1)-\bar{G}(w+1)| \\ \le& \sqrt{1-b}|\tilde{\beta}_n(w+1)-\tilde{\beta}(w+1)| + ab\sqrt{1-b}C_2\epsilon.\end{aligned}$$

Applying this inequality recursively, we have that for all $w \ge w_1$ and $n \ge n_1$,

$$|\tilde{\beta}_n(w)-\tilde{\beta}(w)| \le ab\sqrt{1-b}C_2 e \sum_{i=1}^{\infty}\left(\sqrt{1-b}\right)^{i-1} = \frac{ab\sqrt{1-b}C_2\epsilon}{1-\sqrt{1-b}} = C_3\epsilon, \quad (51)$$

where $C_3 = ab\sqrt{1-b}C_2/(1-\sqrt{1-b})$. It follows from (50) that for $w < w_1$,

$$\begin{aligned}&|\tilde{\beta}_n(w)-\tilde{\beta}(w)| \\ &\quad\le \frac{b^2(1-b)|\beta_n(w+1)-\beta'(w+1)|}{\tilde{\beta}_n(w+1)(1-a\bar{G}(w+1))\tilde{\beta}(w+1)} \\ &\qquad+ \frac{ab^2(1-b)|\bar{G}_n(w+1)-\bar{G}(w+1)|}{(1-a\bar{G}_n(w+1))\tilde{\beta}_n(w+1)(1-a\bar{G}(w+1))} \\ &\quad\le \frac{b^2(1-b)|\tilde{\beta}_n(w+1)-\tilde{\beta}(w+1)|}{(b-a)^2(1-a)} + \frac{ab^2(1-b)C_2\epsilon}{(1-a)^2(b-a)}.\end{aligned} \quad (52)$$

The last inequality follows from $\bar{G}_n(w+1) \le 1$ and $\bar{G}(w+1) \le 1$ and from Lemma 2 that

$$\tilde{\beta}_n(w) \ge \frac{b-a(1-\underline{q})^w}{1-a(1-\underline{q})^w} \ge b-a \quad \text{and} \quad \tilde{\beta}(w) \ge b-a.$$

By applying (52) recursively, we have that for all $w < w_1$ and $n \geq n_1$

$$\begin{aligned} |\tilde{\beta}_n(w) - \tilde{\beta}(w)| &\leq \left(\frac{b^2(1-b)}{(b-a)^2(1-a)}\right)^{w_1-w} |\tilde{\beta}_n(w_1) - \tilde{\beta}(w_1)| \\ &\quad + \frac{ab^2(1-b)C_2\epsilon}{(1-a)^2(b-a)} \sum_{i=0}^{w_1-w-1} \left(\frac{b^2(1-b)}{(b-a)^2(1-a)}\right)^i \\ &\leq \left(\frac{b^2(1-b)}{(b-a)^2(1-a)}\right)^{w_1-w} C_3\epsilon \\ &\quad + \frac{ab^2(1-b)C_2\epsilon}{(1-a)^2(b-a)} \sum_{i=0}^{w_1-w-1} \left(\frac{b^2(1-b)}{(b-a)^2(1-a)}\right)^i \leq c_1\epsilon, \end{aligned} \tag{53}$$

where

$$\begin{aligned} c_1 = \sup_{1 \leq w \leq w_1} &\left(\frac{b^2(1-b)}{(b-a)^2(1-a)}\right)^{w_1-w} C_3 \\ &+ \frac{ab^2(1-b)C_2}{(1-a)^2(b-a)} \sum_{i=0}^{w_1-w-1} \left(\frac{b^2(1-b)}{(b-a)^2(1-a)}\right)^i. \end{aligned}$$

Note that $w_1$ is independent of $\epsilon$. Thus, the constant $c_1$ is independent of $\epsilon$ as well. Combining Eqs. (51) and (53), we have that

$$|\tilde{\beta}_n(w) - \tilde{\beta}(w)| \leq c_2\epsilon, \;\; w \geq 1 \;\text{ and } n \geq n_1,$$

where $c_2 = \max\{C_3, c_1\}$. By letting $\epsilon \to 0$, we have that $\tilde{\beta}(w) \to \tilde{\beta}$ uniformly. This gives the continuity of $\Phi(\Gamma(\cdot))$. □

*Proof of Corollary 3* Since $\beta^*$ is the solution to the fixed point problem $\beta^* = \Phi(\Gamma(\beta^*))$, it is immediate from Lemma 2 that $\beta^*(w)$ is increasing in $w$ and satisfies inequality (12) for all $w \geq 1$. Note that the left-hand side of Eq. (12) converges to $b$ as $w$ goes to infinity. Therefore, $\lim_{w\to\infty} \beta^*(w) = b$. □

*Proof of Lemma 4* Let $e^* = (\beta^*, q^*)$ be an equilibrium. Let $J^*$ be the expected utility associated with $q^*$. It follows from Proposition 2 that $J^* = T_{\beta^*} J^*$, where $T_{\beta^*}$ is given by (44). Let $\mathscr{V} = \{J \in l^\infty : J(w_1) \leq J(w_2) \leq r, 1 \leq w_1 \leq w_2\}$. We first show that for any $J_0 \in \mathscr{V}$, $J = T_{\beta^*} J^0 \in \mathscr{V}$. It follows from (44) that for $w \geq 1$,

$$\begin{aligned} J(w) &= \mathbb{E}_\varepsilon \max\{\varepsilon(1), -c + \alpha[\beta^*(w)r + (1-\beta^*(w))J^0(w+1)] + \varepsilon(0)\} \\ &\leq \mathbb{E}_\varepsilon \max\{\varepsilon(1), -c + \alpha[\beta^*(w+1)r + (1-\beta^*(w+1))J^0(w+1)] + \varepsilon(0)\} \\ &\leq \mathbb{E}_\varepsilon \max\{\varepsilon(1), -c + \alpha[\beta^*(w+1)r + (1-\beta^*(w+1))J^0(w+2)] + \varepsilon(0)\} \\ &= J(w+1). \end{aligned} \tag{54}$$

The first inequality follows from Corollary 3 and the assumption that $J_0 \in \mathscr{V}$. In particular, $\beta^*(w) \leq \beta^*(w+1)$ and $J_0(w+1) \leq r$. The second inequality follows from the assumption that $J^0(w+1) \leq J^0(w+2)$. In addition, the following holds: For $w \geq 1$,

$$\begin{aligned} J(w) &= \mathbb{E}_\varepsilon \max\{\varepsilon(1), -c + \alpha[\beta^*(w)r + (1-\beta^*(w))J^0(w+1)] + \varepsilon(0)\} \\ &\leq \mathbb{E}_\varepsilon \max\{\varepsilon(1), -c + \alpha r + \varepsilon(0)\} \leq r, \end{aligned} \tag{55}$$

where the first inequality follows from $J_0(w+1) \leq r$ and the second inequality follows from Assumption 2. Therefore, it follows from (54) to (55) that $J \in \mathscr{V}$. We have shown in the proof of Proposition 2 that $T$ is a contraction mapping. Since $\mathscr{V}$ is a closed set, we have that $J^* = \lim_{n\to\infty} T^n_{\beta^*} J^0 \in \mathscr{V}$. In particular, $J^*(w)$ is increasing in $w$ and bounded above by $r$.

In addition, it follows from (10) that for $w \geq 1$,

$$\begin{aligned} q^*(w) =& \bar{F}(-c + \alpha[\beta^*(w)r + (1-\beta^*(w))J^*(w+1)]) \\ \geq& \bar{F}(-c + \alpha[\beta^*(w+1)r + (1-\beta^*(w+1))J^*(w+2)]) \\ =& q^*(w+1). \end{aligned}$$

The inequality follows from Corollary 3 that $\beta^*(w) \leq \beta^*(w+1)$ and the monotonicity of $J^*$, i.e., $J^*(w+1) \leq J^*(w+2) \leq r$. Thus, $q^*(w)$ is decreasing in $w$. □

*Proof of Corollary 4* It follows from Lemma 4 and Corollary 1 that $J^*(w)$ is increasing in $w$ and bounded above by $r$. Thus, there exists $J'_\infty \leq r$ such that $\lim_{w\to\infty} J^*(w) = J'_\infty$.

Note that the right-hand side of (14) equals $\kappa(b, x)$, where $\kappa(\cdot)$ is defined in (95). It follows from Lemma 19 that the fixed point of (14) is unique. Let $J_\infty = j(b)$ be the fixed point of (14), where $j(\cdot)$ is given in Lemma 19.

Next, we show that $J'_\infty = J_\infty$. We first show that $J'_\infty \leq J_\infty$. Let $\beta_1(w) = b$ and $J_1(w) = J_\infty$ for all $w \geq 1$. It is immediate that $J_1$ is a fixed point of $J = T_{\beta_1} J$, where the operator $T_{\beta_1}$ is given by (44). As shown in the proof of Proposition 2 that $T_{\beta_1}$ is a contraction mapping, this fixed point is unique. Substituting the inequalities $\beta^*(w) \leq \beta_1(w) = b$ and $J^*(w) \leq r$ into (44), we have that

$$J^*(w) = T_{\beta^*} J^*(w) \leq T_{\beta_1} J^*(w), \ w \geq 1.$$

Substituting this inequality recursively into (44), we have that $J^*(w) \leq T^n_{\beta_1} J^*(w)$ for all $n$, $w$. Thus, the following holds:

$$J^*(w) \leq \lim_{n\to\infty} T^n_{\beta_1} J^*(w) = J_1(w) = J_\infty, \ w \geq 1. \tag{56}$$

Letting $w$ go to infinity, we have that $J'_\infty = \lim_{w\to\infty} J^*(w) \leq J_\infty$.

Next we show that $J'_\infty \geq J_\infty$. It follows from Corollary 3 that $\beta^*(w) \to b$. Thus, fixing $\epsilon > 0$, there exists $w_1$ such that $\beta^*(w) \geq b - \epsilon$ for all $w \geq w_1$. Let $\beta_2(w) = \beta^*(w + w_1)$ and $\beta_3(w) = b - \epsilon$ for $w \geq 1$. In addition, let $J_2(w) = J^*(w + w_1)$ and

$J_3(w) = j(b-\epsilon)$, where $j(\cdot)$ is given in Lemma 19. It is immediate that $J_i$ is the unique fixed point of $J = T_{\beta_i} J$, $i = 2, 3$. Since $\beta_2(w) \geq \beta_3(w)$ for all $w \geq 1$, we can repeat the proof of (56) and show that $J_2(w) \geq J_3(w) = j(b-\epsilon)$ for all $w \geq 1$. Letting $w$ go to infinity, we obtain that

$$J'_\infty = \lim_{w\to\infty} J^*(w) = \lim_{w\to\infty} J_2(w) \geq j(b-\epsilon).$$

By letting $\epsilon \to 0$, it follows from the continuity of $j(\cdot)$ (cf. Lemma 20) that $J'_\infty \geq j(b) = J_\infty$. Thus, we conclude that $\lim_{w\to\infty} J^*(w) = J'_\infty = J_\infty$.

It follows from Lemma 4 and Corollary 1 that $q^*(w)$ is decreasing in $w$ and bounded above from $\underline{q}$. Thus, there exists a constant $q_\infty$ such that $\lim_{w\to\infty} q^*(w) = q_\infty$. In addition, it follows from (10) that

$$\begin{aligned} q_\infty = \lim_{w\to\infty} q^*(w) &= \lim_{w\to\infty} \bar{F}(-c+\alpha[\beta^*(w)r + (1-\beta^*(w))J^*(w+1)]) \\ &= \bar{F}(-c+\alpha(br+(1-b)J_\infty)). \end{aligned}$$

The last inequality follows from the continuity of $\bar{F}(\cdot)$ and that $\beta^*(w) \to b$ and $J^*(w) \to J_\infty$ as $w \to \infty$. □

**Proofs of the Proposition and the Lemma in Section 4**

*Proof of Lemma 9* Fixing $N$ and comparing (37)–(40) and (70)–(72), we have that

$$(e_N(w), \bar{G}_N(w)) = h(e_N(w+1), \bar{G}_N(w+1)), \ w < N. \tag{57}$$

Note that the truncation in (72) is immaterial in this case because $\bar{G}_N(w) \leq 1$ for $w \geq 1$. Fixing $w = N$ and substituting $z(N) = (\beta_N(N), q_N(N), \bar{G}_N(N))$ into Eq. (83), we have that the resulting $z(1) = (\beta_N(1), q_N(1), \bar{G}_N(1))$ satisfies (84). In particular, $\bar{G}_N(1) = 1 - q_N(1)$. Thus, it follows from the definition of the function $f_N(\cdot)$ that

$$\bar{G}_N(w) = f_w(e_N(w)), w \geq N. \tag{58}$$

In particular, $\bar{G}_N(N) = f_N(e_N(N))$. In other words, the value of $\bar{G}_N(N)$ is uniquely determined. Since the truncated equilibrium is fully characterized by $\bar{G}_N(N)$, we conclude that the truncated equilibrium is unique. □

*Proof of Proposition 5* To facilitate the analysis to follow, we define a function $\tilde{h} = (\tilde{h}_1, \tilde{h}_2)$ as follows: For $w \geq 1$ and $(\beta, q) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w) \subseteq (0, b] \times [q_\infty, 1)$,

$$\tilde{h}_i(\beta, q; w) = h(\beta, q, f_w(\beta, q)), \ i = 1, 2,$$

where the functions $h(\cdot)$ and $f_w(\cdot)$ are defined in (70)–(72) and (82)–(84) and $\mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$ is given in (101). Define a matrix $D\tilde{h}(\beta_1, q_1, \beta_2, q_2; w)$ as follows: For $w \geq 1$ and $(\beta_1, q_1), (\beta_2, q_2) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$,

$$D\tilde{h}(\beta_1, q_1, \beta_2, q_2; w) = \begin{bmatrix} Dh_{11}(\beta_1, q_1; w) & Dh_{12}(\beta_1, q_1; w) \\ Dh_{21}(\beta_2, q_2; w) & Dh_{22}(\beta_2, q_2; w) \end{bmatrix},$$

where

$$D\tilde{h}_{11}(\beta, q; w) = \frac{\partial h_1(\beta, q, f_w(\beta, q))}{\partial z_1} + \frac{\partial h_1(\beta, q, f_w(\beta, q))}{\partial z_3} \frac{\partial f_w(\beta, q)}{\partial \beta},$$
$$D\tilde{h}_{12}(\beta, q; w) = \frac{\partial h_1(\beta, q, f_w(\beta, q))}{\partial z_2} + \frac{\partial h_1(\beta, q, f_w(\beta, q))}{\partial z_3} \frac{\partial f_w(\beta, q)}{\partial q},$$
$$D\tilde{h}_{21}(\beta, q; w) = \frac{\partial h_2(\beta, q, f_w(\beta, q))}{\partial z_1} + \frac{\partial h_2(\beta, q, f_w(\beta, q))}{\partial z_3} \frac{\partial f_w(\beta, q)}{\partial \beta},$$
$$D\tilde{h}_{22}(\beta, q; w) = \frac{\partial h_2(\beta, q, f_w(\beta, q))}{\partial z_2} + \frac{\partial h_2(\beta, q, f_w(\beta, q))}{\partial z_3} \frac{\partial f_w(\beta, q)}{\partial q}.$$

It is immediate that $D\tilde{h}(\beta, q, \beta, q; w)$ is the Jacobian matrix of $\tilde{h}(\beta, q)$. In addition, define a constant matrix $Dh_0$ as follows:

$$D\tilde{h}_0 = \begin{bmatrix} \frac{\partial h_1(z_0)}{\partial z_1} & \frac{\partial h_1(z_0)}{\partial z_2} \\ \frac{\partial h_2(z_0)}{\partial z_1} & \frac{\partial h_2(z_0)}{\partial z_2} \end{bmatrix} = \begin{bmatrix} 1-b & 0 \\ -f(\bar{F}^{-1}(q_\infty))\alpha(r - J_\infty)(1-b) & \alpha(1-q_\infty)(1-b) \end{bmatrix},$$

where $z_0 = (\beta, q_\infty, 0)$. It is immediate that the eigenvalues of $D\tilde{h}_0$ are $1-b$ and $\alpha(1-q_\infty)(1-b)$. Thus, there exists an invertible matrix $S$ such that the following holds:

$$\begin{bmatrix} 1-b & 0 \\ 0 & \alpha(1-q_\infty)(1-b) \end{bmatrix} = S(D\tilde{h}_0)S^{-1}.$$

Define a vector norm$||\cdot||_S$ and a matrix norm $|||\cdot|||_S$ as follows: For $x \in \mathbb{R}^2$ and $M \in \mathscr{M}_2$,

$$||x||_S = ||Sx||_\infty \text{ and } |||M|||_S = |||SMS^{-1}|||_\infty.$$

It is immediate that $|||D\tilde{h}_0|||_S = 1-b$. Define a sequence $a_w$ as follows:

$$a_w = \sup\left\{|||D\tilde{h}(\beta_1, q_1, \beta_2, q_2; w)|||_S : (\beta_1, q_1), (\beta_2, q_2) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w)\right\}, \ w \geq 1. \tag{59}$$

We then show that $a_w \to |||D\tilde{h}_0|||_S = 1-b$ as $w \to \infty$. It follows from Lemma 16 that

$$\sup\{|f_w(\beta, q)| : (\beta, q) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w) \subseteq (0, b] \times [q_\infty, 1)\} \to 0 \text{ as } w \to \infty.$$

It follows from Lemma 21 and Eq. (101) that

$$\sup\{|\beta - b| + |q - q_\infty| : (\beta, q) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w)\} \to 0 \text{ as } w \to \infty.$$

Thus, it follows from the continuity of the partial derivatives of $h(\cdot)$ (see (73)–(81)) that for $i = 1, 2$,

$$\sup_{(\beta,q)\in\mathscr{Z}_1(w)\times\mathscr{Z}_2(w)} \left|\frac{\partial h_i(\beta, q, f_w(\beta, q))}{\partial z_j} - \frac{\partial h_i(b, q_\infty, 0)}{\partial z_j}\right| \to 0 \text{ as } w \to \infty. \tag{60}$$

In addition, it follows from Lemma 24 that as $w \to \infty$,

$$\sup_{(\beta,q)\in\mathscr{Z}_1(w)\times\mathscr{Z}_2(w)} \left|\frac{\partial f_w(\beta, q)}{\partial \beta}\right| \to 0 \text{ and } \sup_{(\beta,q)\in\mathscr{Z}_1(w)\times\mathscr{Z}_2(w)} \left|\frac{\partial f_w(\beta, q)}{\partial q}\right| \to 0. \tag{61}$$

Substituting (60)–(61) into $D\tilde{h}(\cdot)$, we have that

$$\sup_{(\beta_1,q_1),(\beta_2,q_2)\in\mathscr{Z}_1(w)\times\mathscr{Z}_2(w)} \left|D\tilde{h}(\beta_1, q_1, \beta_2, q_2; w) - D\tilde{h}_0\right| \to 0 \text{ as } w \to \infty.$$

By the continuity of the norm $|||\cdot|||_S$, we have that

$$\begin{aligned}\lim_{w\to\infty} a_w = &\lim_{w\to\infty} \sup\{|||Dh(\beta_1, q_1, \beta_2, q_2; w)|||_S : (\beta_1, q_1),\\ &(\beta_2, q_2)\in\mathscr{Z}_1(w)\times\mathscr{Z}_2(w)\}\\ =&|||D\tilde{h}_0|||_S = 1 - b.\end{aligned}$$

Thus, there exists $w_1 \geq 1$ such that

$$a_w \leq 1 - b/2 < 1, \; w \geq w_1. \tag{62}$$

Next we show that $e^N \to e^*$ uniformly. Define the difference of the truncated equilibrium and the equilibrium as follows:

$$\delta_\beta^N(w) = \beta_N(w) - \beta^*(w) \text{ and } \delta_q^N(w) = q_N(w) - q^*(w), \; N, w \geq 1.$$

To show that $e^N \to e^*$ uniformly, we need to show that

$$\sup_{w\geq 1} \delta_\beta^N(w) \to 0 \text{ and } \sup_{w\geq 1} \delta_q^N(w) \to 0 \text{ as } N \to \infty.$$

This is equivalent to showing that

$$\sup_{w\geq 1} ||\delta^N(w)||_S \to 0 \text{ as } N \to \infty, \tag{63}$$

where $\delta^N(w) = [\delta_\beta^N(w), \delta_q^N(w)]^T$ for all $N, w \geq 1$. The rest of this proof shows that (63) holds. It follows from Corollaries 3–4 that $\beta^*(w) \to b$ and $q^*(w) \to q_\infty$ as $w \to \infty$. Note that $\beta_N(w) = b$ and $q_N(w) = q_\infty$ for $w \geq N$. Thus, for any $\epsilon > 0$, there exists $N_1 \geq w_1$ such that

$$||\delta^N(w)||_S < \epsilon, \; w \geq N \geq N_1. \tag{64}$$

It follows from (57) to (58), Lemma 25 and Corollary 5 that for $N > w \geq 1$,

$$(\beta_N(w), q_N(w)) = \tilde{h}(\beta_N(w+1), q_N(w+1)) \text{ and}$$
$$(\beta^*(w), q^*(w)) = \tilde{h}(\beta^*(w+1), q^*(w+1)).$$

Thus, it follows from the mean value theorem that for $N > w \geq 1$,

$$\begin{aligned}\delta^N(w) &= \tilde{h}(\beta_N(w+1), q_N(w+1)) - \tilde{h}(\beta_N(w+1), q_N(w+1)) \\ &= D\tilde{h}(w+1; \beta_1^N(w+1), q_1^N(w+1), \beta_2^N(w+1), q_2^N(w+1))\delta^N(w+1),\end{aligned} \tag{65}$$

where

$$\begin{aligned}&(\beta_i^N(w+1), q_i^N(w+1)) \\ &\quad = c_i^N(w+1)(\beta_N(w+1), q_N(w+1)) + (1 - c_i^N(w+1))(\beta^*(w+1), q^*(w+1))\end{aligned}$$

for some $c_i^N(w+1) \in (0, 1)$, $i = 1, 2$. Note that

$$(\beta_N(N), q_N(N)) = (b, q_\infty) \in \mathscr{Z}_1(N) \times \mathscr{Z}_2(N).$$

In addition, it follows from Lemma 26 that $(\beta^*(N), q^*(N)) \in \mathscr{Z}_1(N) \times \mathscr{Z}_2(N)$. It follows from Lemma 22 and the convexity of $\mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$ that

$$(\beta_i^N(w+1), q_i^N(w+1)) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1),\ w = 1, \ldots, N-1, \text{ and } i = 1, 2.$$

Thus, it follows from (59) that for $w = 1, \ldots, N-1$,

$$|||D\tilde{h}(\beta_1^N(w+1), q_1^N(w+1), \beta_2^N(w+1), q_2^N(w+1); w+1)|||_S \leq a_{w+1}.$$

By taking the norm of the both sides of (65), we obtain that, for $w = 1, \ldots, N-1$,

$$\begin{aligned}||\delta^N(w)||_S &\leq |||D\tilde{h}(w+1; \beta_1^N(w+1), q_1^N(w+1), \beta_2^N(w+1), \\ &\quad \times q_2^N(w+1))|||_S ||\delta^N(w+1)||_S \\ &\leq a_{w+1}||\delta^N(w+1)||_S.\end{aligned} \tag{66}$$

Substituting (62) and (64) into (66) yields that for $N \geq N_1$ and $w = w_1, \ldots, N$,

$$||\delta^N(w)||_S \leq \left(1 - \frac{1}{2}b\right)||\delta^N(w+1)||_S \leq \cdots \leq \left(1 - \frac{1}{2}b\right)^{N-w}||\delta^N(N)||_S < \epsilon. \tag{67}$$

In addition, it follows from (66) that for $N \geq N_1$ and $w = 1, \ldots, w_1 - 1$,

$$||\delta^N(w)||_S \leq \left(\prod_{i=w+1}^{w_1} a_i\right)||\delta^N(w_1)||_S < \bar{a}\epsilon,\ w < w_1, \tag{68}$$

where

$$\bar{a} = \max\left\{ \sup_{i\in\{1,\ldots,w_1-1\}} \prod_{j=i+1}^{w_1} a_j, 1 \right\}.$$

Thus, we conclude from (64), (67)–(68) that $||\delta^N(w)||_S < \bar{a}\epsilon$ for all $N \geq N_1$ and $w \geq 1$. By letting $\epsilon \to 0$, Eq. (63) holds. □

*Proof of Lemma 10* We show by induction that for $w = 1, \ldots, N$,

$$\beta_N^1(w) \leq \beta_N^2(w),\;\; q_N^1(w) \geq q_N^2(w) \;\text{ and }\; \bar{G}_N^1(w) > \bar{G}_N^2(w). \tag{69}$$

This is true for $w = N$ by assumption. As the inductive assumption, suppose (69) is true for $w$, then we argue that it is also true for $w-1$. It follows from Eq. (37) and the inductive assumption that $\beta_N^1(w-1) \leq \beta_N^2(w-1)$. Similarly, it follows from (38) to (40) that

$$q_N^1(w-1) \geq q_N^2(w-1) \;\text{ and }\; \bar{G}_N^1(w-1) > \bar{G}_N^2(w-1).$$

In particular, both of the following must be true:

$$q_N^1(1) \geq q_N^2(1) \;\text{ and }\; \bar{G}_N^1(1) > \bar{G}_N^2(1).$$

Thus, the following holds:

$$\bar{G}_N^1(0) = \frac{\bar{G}_N^1(1)}{1-q_N^1(1)} > \bar{G}_N^2(0) = \frac{\bar{G}_N^2(1)}{1-q_N^2(1)}.$$

□

## Appendix 2: Technical lemmas characterizing the equilibrium quantities in discrete time

This section proves Lemma 5 that facilitates the proof of uniqueness of the equilibrium. To prove this result, we define an auxiliary function $f_w(\cdot)$ implicitly and study its properties (especially the monotonicity and convergence of its partial derivatives as $w$ gets large). The function helps characterize $\bar{G}$ in terms of $\beta$ and $q$. We then apply the mean value theorem to $f_w(\cdot)$ to establish the result in Lemma 5.

### Definition of the auxiliary function $f_w(\cdot)$

The function $f_w(\cdot)$ is constructed such that for an equilibrium, the following holds:

$$\bar{G}^*(w) = f_w(\beta^*(w), q^*(w)),\;\; w \geq 1.$$

We establish this relationship in "Characterizing the equilibrium quantities with $f_w(\cdot)$" in Appendix 2. As a preliminary, we first define a function $h(\cdot)$. The function $f_w(\cdot)$ is then defined implicitly as the value satisfying a set of equations characterized by $h(\cdot)$ recursively.

To facilitate the analysis to follow, we define a function $h = (h_1, h_2, h_3) : Z \to \mathbb{R}^3$ as follows[7]:

$$h_1(z) = \left(1 + \frac{1-b}{1-az_3}\frac{1}{z_1}\right)^{-1}, \tag{70}$$

$$h_2(z) = \bar{F}\left(-c + \alpha\left[h_1(z)r + (1-h_1(z))\int_{-\infty}^{\bar{F}^{-1}(z_2)} F(x)\,\mathrm{d}x\right]\right), \tag{71}$$

$$h_3(z) = \min\left(\frac{z_3}{1-z_2}, 1\right), \tag{72}$$

where $z = (z_1, z_2, z_3)$ and $Z = (0, b] \times [q_\infty, 1) \times [0, 1] \subseteq \mathbb{R}^3$, where $q_\infty$ is the constant defined in Corollary 4. The following lemma shows that $h(\cdot)$ maps $Z$ to $Z$.

**Lemma 11** *We have that $h(z) \in Z$ for all $z \in \mathscr{Z}$. In addition, the following inequality holds for all $z \in Z$:*

$$\int_{-\infty}^{\bar{F}^{-1}(z_2)} F(x)\,\mathrm{d}x = \mathbb{E}[\bar{F}^{-1}(z_2) - (\varepsilon(1) - \varepsilon(0))]^+ \le J_\infty \le r,$$

*where $J_\infty$ is the constant defined in Corollary 4.*

*Proof* For any $z \in Z$, it is straightforward that $h_1(z) > 0$. Since $h_1(\cdot)$ is increasing in $z_1$ and decreasing in $z_3$, $h_1(z) \le (1 + (1-b)/b)^{-1} = b$. Thus, $h_1(z) \in (0, b]$.

Recall that the cdf $F(\cdot)$ is the distribution function of the difference of the idiosyncratic shocks $\varepsilon(1) - \varepsilon(0)$. It follows from integration by parts that

$$\begin{aligned}\mathbb{E}[\bar{F}^{-1}(z_2) - (\varepsilon(1) - \varepsilon(0))]^+ &= \int_{-\infty}^{\bar{F}^{-1}(z_2)} \left(\bar{F}^{-1}(z_2) - x\right) \mathrm{d}F(x)\\ &= (\bar{F}^{-1}(z_2) - x)F(x)|_{-\infty}^{\bar{F}^{-1}(z_2)}\\ &\quad - \int_{-\infty}^{\bar{F}^{-1}(z_2)} F(x)\,\mathrm{d}\left(\bar{F}^{-1}(z_2) - x\right)\\ &= \int_{-\infty}^{\bar{F}^{-1}(z_2)} F(x)\,\mathrm{d}x.\end{aligned}$$

[7] We truncate the value of $h_3(z)$ by one to ensure that $h_3(z) \in [0, 1]$ for all $z \in Z$. In the following analysis, we are only interested in $z \in Z$ that satisfies certain conditions. For those $z$ of interest, the truncation is immaterial; see Lemma 15.

By substituting the definitions of $J_\infty$ and $q_\infty$ (in Corollary 4) into the right-hand side, the following inequality holds: For any $z_2 \in [q_\infty, 1)$,

$$\begin{aligned}\int_{-\infty}^{\bar{F}^{-1}(z_2)} F(x)\,\mathrm{d}x &\le \int_{-\infty}^{\bar{F}^{-1}(q_\infty)} F(x)\,\mathrm{d}x \\ &= \mathbb{E}[\bar{F}^{-1}(q_\infty) - (\varepsilon(1) - \varepsilon(0))]^+ \\ &= \mathbb{E}[-c + \alpha[br + (1-b)J_\infty] - (\varepsilon(1) - \varepsilon(0))]^+ \\ &= J_\infty \le r.\end{aligned}$$

The second inequality holds by the definition of $q_\infty$, i.e., $q_\infty = \bar{F}(-c + \alpha[br + (1-b)J_\infty])$. The third equality follows from Eq. (14). This proves that for any $z \in Z$,

$$\int_{-\infty}^{\bar{F}^{-1}(z_2)} F(x) \le J_\infty.$$

Substituting this inequality into (71), we obtain that for $z \in \mathscr{Z}$,

$$\begin{aligned}h_2(z) &= \bar{F}\left(-c + \alpha\left[h_1(z)r + (1-h_1(z))\int_{-\infty}^{\bar{F}^{-1}(z_2)} F(x)\,\mathrm{d}x\right]\right) \\ &\ge \bar{F}\left(-c + \alpha(h_1(z)r + (1-h_1(z))J_\infty)\right) \\ &\ge \bar{F}\left(-c + \alpha(br + (1-b)J_\infty)\right) \\ &= q_\infty,\end{aligned}$$

where the last inequality follows from $h_1(z) \in (0, b]$. In addition, $h_2(z) \le \bar{F}(-c) < 1$. Thus, $h_2(z) \in [q_\infty, 1)$ for any $z \in Z$. Since $z_3 \ge 0$ and $z_2 < 1$, it follows from (72) that $h_3(z) \in [0, 1]$. Thus, $h(z) \in Z$. □

The following lemma shows the elements of the Jacobian matrix of $h(\cdot)$ and the sign of each element.

**Lemma 12** *The partial derivatives of $h(\cdot)$ are given as follows:*

$$\frac{\partial h_1}{\partial z_1} = \frac{h_1^2(z)}{z_1^2}\frac{1-b}{1-az_3} > 0, \tag{73}$$

$$\frac{\partial h_1}{\partial z_2} = 0, \tag{74}$$

$$\frac{\partial h_1}{\partial z_3} = -\frac{ah_1^2(z)}{(1-az_3)^2}\frac{1-b}{z_1} < 0, \tag{75}$$

$$\frac{\partial h_2}{\partial z_1} = -f(\bar{F}^{-1}(h_2(z)))\alpha\left(r - \int_{-\infty}^{\bar{F}^{-1}(z_2)} F(x)\,\mathrm{d}x\right)\frac{h_1^2(z)}{z_1^2}\frac{1-b}{1-az_3} < 0, \tag{76}$$

$$\frac{\partial h_2}{\partial z_2} = \alpha(1-h_1(z))(1-z_2)\frac{f(\bar{F}^{-1}(h_2(z)))}{f(\bar{F}^{-1}(z_2))} > 0, \tag{77}$$

$$\frac{\partial h_2}{\partial z_3} = f(\bar{F}^{-1}(h_2(z))\alpha \left( r - \int_{-\infty}^{\bar{F}^{-1}(z_2)} F(x)\,\mathrm{d}x \right) \frac{a h_1^2(z)}{(1-az_3)^2} \frac{1-b}{z_1} > 0, \tag{78}$$

$$\frac{\partial h_3}{\partial z_1} = 0, \tag{79}$$

$$\frac{\partial h_3}{\partial z_2} = \begin{cases} \dfrac{z_3}{(1-z_2)^2} \geq 0, & \textit{if } z_3 < 1 - z_2, \\ 0, & \textit{if } z_3 > 1 - z_2, \end{cases} \tag{80}$$

$$\frac{\partial h_3}{\partial z_3} = \begin{cases} \dfrac{1}{1-z_2} > 1, & \textit{if } z_3 < 1 - z_2, \\ 0, & \textit{if } z_3 > 1 - z_2. \end{cases} \tag{81}$$

*Proof* Since $h(\cdot)$ is given explicitly by Eqs. (70)–(72), the partial derivatives of $h(\cdot)$ are immediate. The signs of Eqs. (76) and (78) follow from Lemma 11. To be specific,

$$r - \int_{-\infty}^{\bar{F}^{-1}(z_2)} F(x)\,\mathrm{d}x \geq 0, \;\; z_2 \in [q_\infty, 1).$$

The signs of other equations are immediate.

In addition, the following lemma will be useful in the analysis to follow.

**Lemma 13** *We have that* $h_3(z) \geq z_3$ *for all* $z \in Z$.

*Proof* For all $z \in Z$, the following inequality holds:

$$h_3(z) = \min\left( \frac{z_3}{1-z_2}, 1 \right) \geq \min(z_3, 1) = z_3.$$

For every $w \geq 1$, we define an implicit function $f_w(\beta, q) : (0, b] \times [q_\infty, 1) \to [0, 1]$ through the set of equations immediately below. Namely, for $(\beta, q) \in (0, b] \times [q_\infty, 1)$, $f_w(\beta, q)$ and $z(k)$ for $k = w, \ldots, 1$ are defined implicitly by Eqs. (82)–(84).

$$z(w) = (\beta, q, f_w(\beta, q)), \tag{82}$$

$$z(k-1) = h(z(k)) \;\text{ for }\; k = w, \ldots, 2, \tag{83}$$

$$z_2(1) = 1 - z_3(1). \tag{84}$$

Intuitively, given $(\beta, q)$ and an initial guess of $f_w(\beta, q)$, $z(w)$ is defined by (82) and $z(k)$ (for $k = 1, \ldots, w-1$) are well-defined by (83) and Lemma 11. The essence of what the next lemma shows is that there is a unique value of $f_w(\cdot)$ such that the boundary condition (84) is satisfied.

**Lemma 14** *The function* $f_w(\cdot)$ *is well-defined for all* $w \geq 1$. *In addition,* $f_w(\beta, q) \in (0, 1)$ *for all* $(\beta, q) \in (0, b] \times [q_\infty, 1)$.

*Proof* Note that given $z(w) = (\beta, q, f_w(\beta, q))$, Eq. (83) defines $z(k)$ for $k = 1, \ldots, w-1$. However, the resulting $z$ values must also satisfy the boundary condition (84). In other words, we need to show that for any $(\beta, q) \in (0, b] \times [q_\infty, 1)$, there exists a unique value $f_w(\beta, q)$ such that the resulting $z(k), k = 1, \ldots, w$, satisfy (82)–(84).

We first show that there exists a value of $f_w(\beta, q) = \eta$ such that (82)–(84) are satisfied. To this end, we view $z(k)$ for $k = 1, \ldots, w-1$ as functions of $\eta$, denoted by $z(k; \eta)$ for $\eta \in [0, 1]$. In addition, we define a function $\phi(\eta)$ as follows:

$$\phi(\eta) = z_2(1; \eta) - (1 - z_3(1; \eta)).$$

Note that $z_3(1; 0) = 0$ which follows from Eq. (72) inductively. It follows from Lemma 11 by induction that $z_2(1; \eta) \in [q_\infty, 1)$ for all $\eta$. In particular, $z_2(1; 0) < 1$, so $\phi(0) = z_2(1; 0) - 1 < 0$. Next, we argue that $\phi(1) > 0$. To see this, note that for any $\eta \in [0, 1]$,

$$1 \geq z_3(1; \eta) \geq z_3(2; \eta) \geq \cdots \geq z_3(w; \eta) = \eta \geq 0, \tag{85}$$

which follows from Lemma 13 inductively. Note from Eq. (85) that $z_3(1; 1) = 1$.

Thus, $\phi(1) = z_2(1; 1) \geq q_\infty > 0$.

Since $h(\cdot)$ is continuous, $\phi(\cdot)$ is continuous as well. Combining the facts that $\phi(0) < 0$ and $\phi(1) > 0$, we conclude that there exists $\eta \in (0, 1)$ such that $\phi(\eta) = 0$, i.e., condition (84) is satisfied. For such $\eta$, the resulting vectors $z(k; \eta)$, $k = 1, \ldots, w$ satisfy conditions (82)–(84).

For $\eta$ such that $\phi(\eta) = 0$, $z_3(1; \eta) = 1 - z_2(1; \eta) < 1$. It follows from Eq. (85) that $z_3(k; \eta) < 1$. Substituting the definition of $h_3(\cdot)$ into the inequality $z_3(k; \eta) < 1$, we argue that for such $\eta$, the following holds:

$$z_3(k; \eta) = \frac{z_3(k+1; \eta)}{1 - z_2(k+1; \eta)}, \quad k = 1, \ldots, w, \tag{86}$$

because $z_3(k; \eta)$ cannot take the value 1. In other words, the truncation by 1 on the right-hand side of (72) is immaterial for solutions of $f_w(\beta, q)$ and $z(k)$ for $k = 1, \ldots, w-1$ defined through (82)–(84).

We conclude the proof by showing that there exists a unique $\eta$ satisfying conditions (82)–(84). Suppose there are multiple values of $f_w(\beta, q)$, say $\eta \neq \tilde{\eta}$, satisfying the conditions (82)–(84). Without loss of generality, assume $\eta > \tilde{\eta}$. Next, we show by induction that for $k = 1, \ldots, w$,

$$z_1(k; \eta) \leq z_1(k; \tilde{\eta}), \quad z_2(k; \eta) \geq z_2(k; \tilde{\eta}) \text{ and } z_3(k; \eta) > z_3(k; \tilde{\eta}). \tag{87}$$

This is true for $k = w$ by assumption. Suppose it is true for $k$, then we argue that it is also true for $k-1$. Note that $z_1(k-1; \eta) = h_1(z(k; \eta))$ and $z_1(k-1; \tilde{\eta}) = h_1(z(k; \tilde{\eta}))$. Since $h_1(\cdot)$ is decreasing in its first argument, whereas increasing in its last two arguments, we conclude that $z_1(k-1; \eta) \leq z_1(k-1; \tilde{\eta})$. Similarly, because

$h_2(\cdot)$ is decreasing in the first argument, whereas increasing in its last two arguments, we conclude that

$$z_2(k-1;\eta) = h_2(z(k;\eta)) \geq h_2(z(k;\tilde{\eta})) = z_2(k-1;\tilde{\eta}).$$

Also, it follows from Eq. (86) that

$$z_3(k-1;\eta) = \frac{z_3(k;\eta)}{1-z_2(k;\eta)} > \frac{z_3(k;\tilde{\eta})}{1-z_2(k;\tilde{\eta})} = z_3(k-1;\tilde{\eta}).$$

In particular, both of the following must be true:

$$z_3(1;\eta) > z_3(1;\tilde{\eta}) \text{ and } z_2(1;\eta) \geq z_2(1;\tilde{\eta}).$$

However, by Eq. (84), we conclude that

$$z_2(1;\eta) = 1 - z_3(1;\eta) < 1 - z_3(1;\tilde{\eta}) = \tilde{z}_2(1;\tilde{\eta}),$$

which is a contraction. Therefore, there exists at most one value of $f_w(\beta, q)$ satisfying conditions (82)–(84).

### Partial derivatives of the auxiliary function $f_w(\cdot)$

This subsection characterizes the partial derivatives of $f_w(\cdot)$ and establishes the monotonicity of $f_w(\cdot)$.

To facilitate the analysis to follow, fix $w \geq 1$ and denote by $z(k; w, \beta, q)$ (for $k = 1, \ldots, w$) the $z(k)$ defined by substituting $z(w) = (\beta, q, f_w(\beta, q))$ into Eq. (83). The following lemma shows that in this construction, the truncation by 1 in defining $h_3(\cdot)$ is immaterial, cf. Eq. (72).

**Lemma 15** *For* $w \geq 1$ *and* $(\beta, q) \in (0, b] \times [q_\infty, 1)$*, the following holds:*

$$z_3(k; w, \beta, q) = h_3(z(k+1; w, \beta, q)) = \frac{z_3(k+1; w, \beta, q)}{1 - z_2(k+1; w, \beta, q)}, \quad k = 1, \ldots, w-1.$$

*Proof* It follows from Lemma 13 inductively that

$$z_3(1; w, \beta, q) \geq z_3(2; w, \beta, q) \geq \cdots \geq z_3(w; w, \beta, q) = f_w(\beta, q) > 0.$$

In addition, condition (84) ensures that $z_3(1; w, \beta, q) = 1 - z_2(1; w, \beta, q) \leq 1 - q_\infty < 1$. Combining these inequalities with Eq. (72) yields the following: For $k = 1, \ldots, w-1$,

$$1 > z_3(k; w, \beta, q) = h_3(z(k+1; w, \beta, q)) = \min\left(1, \frac{z_3(k+1; w, \beta, q)}{1 - z_2(k+1; w, \beta, q)}\right).$$

Since $z_3(k; w, \beta, q)$ cannot take the value 1, the result follows.

The following lemma provides an upper bound of $z_3(k; w, \beta, q)$ for $w \geq 1$.

**Lemma 16** *For* $w \geq 1$ *and* $(\beta, q) \in (0, b] \times [q_\infty, 1)$*, we have the following inequality:*

$$z_3(k; w, \beta, q) \leq (1 - q_\infty)^k, \; k = 1, \ldots, w.$$

*In particular,* $f_w(\beta, q) \leq (1 - q_\infty)^w$ *for* $w \geq 1$.

*Proof* We proceed by induction. On the induction basis, it follows from Eq. (84) that for $k = 1$,

$$z_3(1; w, \beta, q) = (1 - z_2(1; w, \beta, q)) \leq 1 - q_\infty,$$

where the inequality follows because $z_2(1; w, \beta, q) \in [q_\infty, 1)$ by construction.

By the induction hypothesis, suppose that the statement is true for $k$. Then note from Lemma 15 that

$$z_3(k+1; w, \beta, q) = z_3(k; w, \beta, q)(1 - z_2(k+1; w, \beta, q)) \leq (1 - q_\infty)^{k+1},$$

where the inequality follows from the induction hypothesis and that $z_2(k + 1; w, \beta, q) \in [q_\infty, 1)$.

In addition, it follows from (82) that

$$f_w(\beta, q) = z_3(w; w, \beta, q) \leq (1 - q_\infty)^w.$$

The following lemma characterizes the partial derivatives of $f_w(\cdot)$ recursively.

**Lemma 17** *The partial derivatives of* $f_w(\cdot)$ *with respect to* $\beta$ *and* $q$ *for* $w \geq 1$ *and* $(\beta, q) \in (0, b] \times [q_\infty, 1)$ *are given as follows: For* $w = 1$*, we have that*

$$\frac{\partial f_1(\beta, q)}{\partial \beta} = 0 \;\; \text{and} \;\; \frac{\partial f_1(\beta, q)}{\partial q} = -1. \tag{88}$$

*In addition, we have the following recursive characterization for* $w \geq 1$*:*

$$\frac{\partial f_{w+1}(\beta, q)}{\partial \beta} = -\frac{\frac{\partial f_w(e(w))}{\partial \beta}\frac{\partial h_1(z(w+1))}{\partial z_1} + \frac{\partial f_w(e(w))}{\partial q}\frac{\partial h_2(z(w+1))}{\partial z_1}}{\frac{\partial f_w(e(w))}{\partial \beta}\frac{\partial h_1(z(w+1))}{\partial z_3} + \frac{\partial f_w(e(w))}{\partial q}\frac{\partial h_2(z(w+1))}{\partial z_3} - \frac{\partial h_3(z(w+1))}{\partial z_3}}, \tag{89}$$

$$\frac{\partial f_{w+1}(\beta, q)}{\partial q} = -\frac{\frac{\partial f_w(e(w))}{\partial q}\frac{\partial h_2(z(w+1))}{\partial z_2} - \frac{\partial h_3(z(w+1))}{\partial z_2}}{\frac{\partial f_w(e(w))}{\partial \beta}\frac{\partial h_1(z(w+1))}{\partial z_3} + \frac{\partial f_w(e(w))}{\partial q}\frac{\partial h_2(z(w+1))}{\partial z_3} - \frac{\partial h_3(z(w+1))}{\partial z_3}}, \tag{90}$$

*where* $z(k) = z(k; w+1, \beta, q)$ *for* $k = w, w+1$*, and* $e(w) = (z_1(w), z_2(w))$.

*Proof* Note that $f_1(\beta, q) = 1 - q$. Hence, Eq. (88) is immediate.

Fixing $w \geq 1$ and $(\beta, q) \in (0, b] \times [q_\infty, 1)$, we want to characterize the partial derivatives of $f_{w+1}(\cdot)$ with respect to $\beta$ and $q$. Recall that for $w+1$, $z(w; w+1, \beta, q)$ is computed by substituting $z(w+1) = (\beta, q, f_{w+1}(\beta, q))$ into Eq. (83). In particular, we rewrite Eq. (83) that derives $z(w; w+1, \beta, q)$ as follows:

$$z_i(w; w+1, \beta, q) = h_i(\beta, q, f_{w+1}(\beta, q)). \tag{91}$$

Moreover, by substituting $z(w) = z(w; w+1, \beta, q)$ into Eq. (83) for $w$, we find that condition (84) (for $w$) is satisfied. In other words, solutions of (82)–(84) for different $w$'s are consistent provided that $\beta, q$'s are chosen consistently for each $w$. In particular, the following holds:

$$f_w(z_1(w; w+1, \beta, q), z_2(w; w+1, \beta, q)) = z_3(w; w+1, \beta, q). \tag{92}$$

Substituting Eq. (91) into (92), we obtain the following identity:

$$f_w(h_1(\beta, q, f_{w+1}(\beta, q)), h_2(\beta, q, f_{w+1}(\beta, q))) = h_3(\beta, q, f_{w+1}(\beta, q)). \tag{93}$$

Both the left-hand side and the right-hand side of Eq. (93) are functions of $(\beta, q)$. Since we focus our analysis on the derivation for $w+1$ with fixed initial values $(\beta, q)$, we write $z(k) = z(k; w+1, \beta, q)$ in short.

First, we take the partial derivative of both sides of Eq. (93) with respect to $\beta$ by the chain rule and evaluate the function at point $(\beta, q)$. It follows from Eq. (92) that the partial derivatives of $f_w(\cdot)$ are evaluated at $(z_1(w), z_2(w))$. Since $z(w+1) = (\beta, q, f_{w+1}(\beta, q))$, the partial derivatives of $h_i(\cdot)$ are evaluated at $z(w+1)$ for $i = 1, 2, 3$ . Thus, we obtain the following equation:

$$\begin{aligned}
&\frac{\partial f_w(e(w))}{\partial \beta}\frac{\partial h_1(z(w+1))}{\partial z_1} + \frac{\partial f_w(e(w))}{\partial \beta}\frac{\partial h_1(z(w+1))}{\partial z_3}\frac{\partial f_{w+1}(\beta, q)}{\partial \beta}\\
&\quad + \frac{\partial f_w(e(w))}{\partial q}\frac{\partial h_2(z(w+1))}{\partial z_1} + \frac{\partial f_w(e(w))}{\partial q}\frac{\partial h_2(z(w+1))}{\partial z_3}\frac{\partial f_{w+1}(\beta, q)}{\partial \beta}\\
&\quad = \frac{\partial h_3(z(w+1))}{\partial z_1} + \frac{\partial h_3(z(w+1))}{\partial z_3}\frac{\partial f_{w+1}(\beta, q)}{\partial \beta},
\end{aligned}$$

where $e(w) = (z_1(w), z_2(w))$. Note that $\partial h_3 / \partial z_1 = 0$ by (74). Thus, we can drop the first term on the right-hand side. Rearranging the terms yields Eq. (89).

Taking the partial derivative of both sides of Eq. (93) with respect to $q$ and evaluating the function at value $(\beta, q)$, we obtain the following equation:

$$\begin{aligned}
&\frac{\partial f_w(e(w))}{\partial \beta}\frac{\partial h_1(z(w+1))}{\partial z_2} + \frac{\partial f_w(e(w))}{\partial \beta}\frac{\partial h_1(z(w+1))}{\partial z_3}\frac{\partial f_{w+1}(\beta, q)}{\partial q}\\
&\quad + \frac{\partial f_w(e(w))}{\partial q}\frac{\partial h_2(z(w+1))}{\partial z_2} + \frac{\partial f_w(e(w))}{\partial q}\frac{\partial h_2(z(w+1))}{\partial z_3}\frac{\partial f_{w+1}(\beta, q)}{\partial q}\\
&\quad = \frac{\partial h_3(z(w+1))}{\partial z_2} + \frac{\partial h_3(z(w+1))}{\partial z_3}\frac{\partial f_{w+1}(\beta, q)}{\partial q}.
\end{aligned}$$

It follows from (79) that $\partial h_1/\partial z_2 = 0$. Thus, we can drop the first term on the left-hand side. Rearranging the terms yields Eq. (90).

The following lemma shows the monotonicity of $f_w(\cdot)$.

**Lemma 18** *$f_w$ is non-decreasing in $\beta$ and non-increasing in $q$. That is, for all $w \geq 1$ and $(\beta, q) \in (0, b] \times [q_\infty, 1)$,*

$$\frac{\partial f_w(\beta, q)}{\partial \beta} \geq 0 \;\; \textit{and} \;\; \frac{\partial f_w(\beta, q)}{\partial q} \leq 0. \tag{94}$$

*Proof* Recall that $f_1(\beta, q) = 1 - q$. Thus, (94) is immediate for $w = 1$.

We proceed by induction: Suppose (94) holds for $k = 1, \ldots, w$, and we next show that it holds for $k = w + 1$. Note from Eqs. (73) to (81) that

$$\frac{\partial h_1}{\partial z_1}, \; \frac{\partial h_2}{\partial z_3}, \; \frac{\partial h_3}{\partial z_3} > 0 \;\; \text{and} \;\; \frac{\partial h_2}{\partial z_1}, \; \frac{\partial h_1}{\partial z_3} < 0.$$

Consider the formula for $\partial f_{w+1}/\partial \beta$ given in Eq. (89). Every term in the numerator is positive, whereas every term in the denominator is negative so that

$$\frac{\partial f_{w+1}(\beta, q)}{\partial \beta} \geq 0.$$

Next, consider $\partial f_{w+1}/\partial q$. It follows from Eqs. (73) to (81) that

$$\frac{\partial h_2}{\partial z_2} > 0 \;\; \text{and} \;\; \frac{\partial h_3}{\partial z_2} \geq 0.$$

Every term in both the numerator and the denominator of Eq. (90) is negative. Thus, we conclude that

$$\frac{\partial f_{w+1}(\beta, q)}{\partial q} \leq 0.$$

**Properties of $f_w(\cdot)$ on a restricted set**

This subsection studies the partial derivatives of $f_w(\beta, q)$ as $w$ gets large. To facilitate this analysis, we define subsets $\mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$ for $w \geq 1$ such that for any potential equilibrium $(\beta^*(w), q^*(w)) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$ for all $w \geq 1$. Restricting our analysis to the case where $(\beta, q) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w) \subseteq (0, b] \times [q_\infty, 1)$, we establish the desired convergence results for the partial derivatives of $f_w(\beta, q)$. (Note from Corollaries 3 and 4 that $(\beta^*(w), q^*(w)) \to (b, q_\infty)$ as $w \to \infty$ for any potential equilibrium. We define $\mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$ such that they shrink to the point $(b, q_\infty)$ as $w \to \infty$. )

To facilitate the analysis to follow, we first define a function $\kappa(x, y) : [0, 1] \times [0, r] \to \mathbb{R}$ as follows:

$$\kappa(x, y) = \mathbb{E}[-c + \alpha[xr + (1 - x)y] - (\varepsilon(1) - \varepsilon(0))]^+. \tag{95}$$

The following lemma shows the properties of $\kappa$.

**Lemma 19** *The function $\kappa(x, y)$ has the following properties:*

*(i) For any $(x, y) \in [0, 1] \times [0, r]$, $\kappa(x, y) \in [0, r)$.*
*(ii) For any fixed $x \in [0, 1]$, $\kappa(x, y)$ is a contraction mapping. In particular, $|\partial\kappa(x, y)/\partial y| \leq \alpha < 1$ for all $x \in [0, 1]$.*
*(iii) For any fixed $x \in [0, 1]$, there exists a unique $j(x) \in [0, r)$ satisfying $j(x) = \kappa(x, j(x))$.*

*Proof* The following inequality shows that (i) holds: For any $(x, y) \in [0, 1] \times [0, r]$,

$$\begin{aligned} \kappa(x, y) &= \mathbb{E}[-c + \alpha[xr + (1 - x)y] - (\varepsilon(1) - \varepsilon(0))]^+ \\ &\leq \mathbb{E}[-c + \alpha[xr + (1 - x)r] - (\varepsilon(1) - \varepsilon(0))]^+ \\ &= \mathbb{E}[-c + \alpha r - (\varepsilon(1) - \varepsilon(0))]^+ \\ &= \mathbb{E}[\max\{\varepsilon(1), -c + \alpha r + \varepsilon(0)\}] < r. \end{aligned} \tag{96}$$

The first inequality follows from $y \leq r$. The equality in the fourth line holds because $\mathbb{E}[\varepsilon(1)] = 0$. The last equality follows from Assumption 2. In addition, it is immediate that $\kappa(x, y) \geq 0$ for all $(x, y) \in [0, 1] \times [0, r]$. Thus, we have that $\kappa(x, y) \in [0, r)$.

We can write $\kappa(x, y)$ in integral form and use integration by parts to arrive at the following:

$$\begin{aligned} \kappa(x, y) &= \mathbb{E}[-c + \alpha[xr + (1 - x)y] - (\varepsilon(1) - \varepsilon(0))]^+ \\ &= \int_{-\infty}^{-c+\alpha[xr+(1-x)y]} (-c + \alpha[xr + (1 - x)y] - u)\, \mathrm{d}F(u) \\ &= \int_{-\infty}^{-c+\alpha[xr+(1-x)y]} F(u)\, \mathrm{d}u, \end{aligned} \tag{97}$$

where the last inequality follows from integration by parts. Thus, the partial derivative of $\kappa(x, y)$ with respect to $y$ is given as follows:

$$\begin{aligned} \frac{\partial\kappa(x, y)}{\partial y} &= \frac{\partial(\int_{-\infty}^{-c+\alpha[xr+(1-x)y]} F(u)\, \mathrm{d}u)}{\partial y} \\ &= \alpha(1 - x)F(-c + \alpha[xr + (1 - x)y]) \in [0, \alpha]. \end{aligned}$$

Therefore, for any fixed $x \in [0, 1]$, the following inequality holds:

$$\left|\frac{\partial\kappa(x, y)}{\partial y}\right| \leq \alpha < 1, \;\; y \in [0, r].$$

Thus, (ii) holds, i.e., $\kappa(x, y)$ is a contraction mapping for any fixed $x \in [0, 1]$.

It follows from properties (i)–(ii) that for any fixed $x \in [0, 1]$, $\kappa(x, y)$ is a contraction mapping from $[0, r]$ to $[0, r)$. By the Banach fixed point theorem, there exists a unique fixed point $j(x) \in [0, r]$ such that $j(x) = \kappa(x, j(x))$. It follows from (96) that

$\kappa(x, r) < r$. Thus, $j(x) \neq r$, which leads to $j(x) \in [0, r)$. In other words, property (iii) holds.

The following lemma shows useful properties of the function $j(\cdot)$.

**Lemma 20** *The function $j(\cdot)$ is increasing and differentiable.*

*Proof* The function $j(x)$ is defined implicitly as follows:

$$\kappa(x, j(x)) - j(x) = 0, \ x \in [0, 1]. \tag{98}$$

Note that $\kappa$ is continuously differentiable by Eq. (97). By the implicit function theorem, $j(x)$ is differentiable; see Theorem 9.28 of [36]. It follows from (97) that

$$\begin{aligned}\frac{\partial \kappa(x, y)}{\partial x} =& \alpha(r - y)F(-c + \alpha[xr + (1 - x)y]) \\ \geq& \alpha(r - y)F(-c) > 0, \ (x, y) \in [0, 1] \times [0, r),\end{aligned}$$

where the last inequality holds because $y < r$ and $-c$ is in the interior of the support of $F(\cdot)$ by Assumption 1. For any fixed $x \in [0, 1]$, taking the derivative of both sides of the equation $j(x) = \kappa(x, j(x))$ yields the following equation:

$$j'(x) = \frac{\partial \kappa(x, j(x))}{\partial x} + \frac{\partial \kappa(x, j(x))}{\partial y} j'(x).$$

Rearranging the terms, we have that

$$j'(x) = \frac{\partial \kappa(x, j(x))}{\partial x} \Bigg/ \left(1 - \frac{\partial \kappa(x, j(x))}{\partial y}\right) > 0, \ x \in [0, 1],$$

where the inequality follows from the fact that $\partial \kappa(x, y)/\partial x > 0$ and property (ii) in Lemma 19. Therefore, $j(x)$ is increasing.

To facilitate the definition of $\mathscr{Z}_1$ and $\mathscr{Z}_2$, the sequence $\underline{\beta}(w)$ is defined as follows:

$$\underline{\beta}(w) = \frac{b - a(1 - q_\infty)^w}{1 - a(1 - q_\infty)^w}, \ w \geq 1.$$

Since $b > a$, we have that $\underline{\beta}(w) > 0$ for all $w \geq 1$. Then we define $\underline{J}(w) = j(\underline{\beta}(w))$. By substituting Eqs. (95) and (98) into the definition of $\underline{J}(w)$, we have that

$$\underline{J}(w) = \mathbb{E}\left[-c + \alpha[\underline{\beta}(w)r + (1 - \underline{\beta}(w))\underline{J}(w)] - (\varepsilon(1) - \varepsilon(0))\right]^+, \ w \geq 1. \tag{99}$$

In addition, define[8]

$$\bar{q}(w) = \bar{F}(-c + \alpha[\underline{\beta}(w)r + (1 - \underline{\beta}(w))\underline{J}(w)]), \ w \geq 1. \tag{100}$$

[8] The sequence $\underline{\beta}(w)$ provides a lower bound of $\beta^*(w)$, while $\bar{q}(w)$ is an upper bound of $q^*(w)$ in any potential equilibrium.

The following lemma shows the properties of the sequences $\underline{\beta}(w)$, $\underline{J}(w)$ and $\bar{q}(w)$ for $w \geq 1$.

**Lemma 21** *The sequences $\underline{\beta}(w)$, $\underline{J}(w)$ and $\bar{q}(w)$ for $w \geq 1$ have the following properties:*

- *(i) $\underline{\beta}(w)$ and $\underline{J}(w)$ are increasing, whereas $\bar{q}(w)$ is decreasing in $w$.*
- *(ii) $\lim_{w\to\infty} \underline{\beta}(w) = b$, $\lim_{w\to\infty} \underline{J}(w) = J_\infty$ and $\lim_{w\to\infty} \bar{q}(w) = q_\infty$, where $J_\infty$ and $q_\infty$ are constants defined in Corollary 4.*
- *(iii) $\underline{J}(w) = \int_{-\infty}^{\bar{F}^{-1}(\bar{q}(w))} F(x)\,\mathrm{d}x$.*

*Proof* We first show (i). It is immediate that $\underline{\beta}(w)$ is increasing. Thus, $\underline{J}(w)$ is increasing in $w$ by Lemma 20. It follows from property (iii) of Lemma 19 that $j(x) \in [0, r)$ for all $x \in [0, 1]$. Thus, $\underline{J}(w) < r$ for $w \geq 1$. It follows from Eq. (100) that $\bar{q}(w)$ is decreasing in $w$ because $\underline{\beta}(w)$ and $\underline{J}(w)$ are non-increasing in $w$ and $\underline{J}(w) < r$.

Next we show that (ii) holds. Clearly, $\lim_{w\to\infty} \underline{\beta}(w) = b$. It follows from Lemma 19 that $j(x)$ is a differentiable function and thus is continuous. Therefore, the following equation holds:

$$\lim_{w\to\infty} \underline{J}(w) = \lim_{w\to\infty} j\left(\underline{\beta}(w)\right) = j\left(\lim_{w\to\infty} \underline{\beta}(w)\right) = j(b) = J_\infty,$$

where the last inequality follows from (14) that $J_\infty = j(b)$. It follows from the continuity of $\bar{F}$ that

$$\begin{aligned}\lim_{w\to\infty} \bar{q}(w) = &\lim_{w\to\infty} \bar{F}\left(-c + \alpha[\underline{\beta}(w)r + (1 - \underline{\beta}(w))\underline{J}(w)]\right)\\ =&\bar{F}\left(-c + \alpha[br + (1 - b)J_\infty]\right) = q_\infty.\end{aligned}$$

The last equality follows from the definition of $q_\infty$ in Corollary 4.

Lastly, we show that (iii) holds. It follows from Eqs. (99) to (100) that for $w \geq 1$,

$$\begin{aligned}\underline{J}(w) =&\mathbb{E}\left[-c + \alpha[\underline{\beta}(w)r + (1 - \underline{\beta}(w))\underline{J}(w)] - (\varepsilon(1) - \varepsilon(0))\right]^+\\ =&\mathbb{E}\left[\bar{F}^{-1}(\bar{q}(w)) - (\varepsilon(1) - \varepsilon(0))\right]^+\\ =&\int_{-\infty}^{\bar{F}^{-1}(\bar{q}(w))} F(x)\,\mathrm{d}x.\end{aligned}$$

To facilitate the analysis, define

$$\mathscr{Z}_1(w) = [\underline{\beta}(w), b] \text{ and } \mathscr{Z}_2(w) = [q_\infty, \bar{q}(w)],\ w \geq 1. \tag{101}$$

It follows from properties (i)–(ii) of Lemma 21 that $\underline{\beta}(w) < b$ and $\bar{q}(w) \geq q_\infty$. Thus, both $\mathscr{Z}_1(w)$ and $\mathscr{Z}_2(w)$ are non-empty for all $w \geq 1$. Since we only consider the underloaded case, i.e., $b > a$, we have that $\underline{\beta}(w) > 0$ for all $w \geq 1$. In addition, it follows from (100) that $\bar{q}(w) < \bar{F}(r) \leq 1$, which gives that

$$\mathscr{Z}_1(w) \times \mathscr{Z}_2(w) \subseteq (0, b] \times [q_\infty, 1),\ \ w \geq 1.$$

Next, we study the properties of $f_w(\beta, q)$ when $(\beta, q) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$.

**Lemma 22** *For any $w \geq 1$, if $(\beta, q) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$, then*

$$(z_1(k; w, \beta, q), z_2(k; w, \beta, q)) \in \mathscr{Z}_1(k) \times \mathscr{Z}_2(k),\ \ k = 1, \ldots, w. \tag{102}$$

*Proof* Fix $w \geq 1$. We proceed by induction. For $k = w$, (102) holds by assumption.

By the induction hypothesis, suppose (102) holds for $l = k+1, \ldots, w-1, w$. That is,

$$\underline{\beta}(l) \leq z_1(l; w, \beta, q) \leq b \ \text{ and } \ q_\infty \leq z_2(l; w, \beta, q) \leq \bar{q}(l),\ \ l = k+1, \ldots, w.$$

Next, we show that (102) holds for $l = k$. The following holds:

$$\begin{aligned}
z_1(k; w, \beta, q) &= h_1(z(k+1; w, \beta, q)) \\
&= \left(1 + \frac{1-b}{1 - a z_3(k+1; w, \beta, q)} \frac{1}{z_2(k+1; w, \beta, q)}\right)^{-1} \\
&> \left(1 + \frac{1-b}{1 - a(1-q_\infty)^{k+1}} \frac{1 - a(1-q_\infty)^{k+1}}{b - a(1-q_\infty)^{k+1}}\right)^{-1} \\
&= \left(1 + \frac{1-b}{b - a(1-q_\infty)^{k+1}}\right)^{-1} \\
&> \left(1 + \frac{1-b}{b - a(1-q_\infty)^{k}}\right)^{-1} \\
&= \frac{b - a(1-q_\infty)^k}{1 - a(1-q_\infty)^k} = \underline{\beta}(k).
\end{aligned}$$

The first inequality follows from Lemma 16 and that $z_2(k+1; w, \beta, q) > \beta(k+1)$. Thus, $z_1(k; w, \beta, q) \in \mathscr{Z}_1(k)$. Combining the two cases, $z_1(k; w, \beta, q) \in \mathscr{Z}_1(k)$.

Moreover, it follows from Eq. (71) that

$$\begin{aligned}
z_2(k; w, \beta, q) &= h_2(z(k+1; w, \beta, q)) \\
&= \bar{F}\left[-c + \alpha\left(z_1(k; w, \beta, q) r + (1 - z_1(k; w, \beta, q))\right.\right. \\
&\qquad \left.\left. \int_{-\infty}^{\bar{F}^{-1}(z_2(k+1; w, \beta, q))} F(x)\,\mathrm{d}x\right)\right] \\
&\leq \bar{F}\left[-c + \alpha\left(\underline{\beta}(k) r + (1 - \underline{\beta}(k)) \int_{-\infty}^{\bar{F}^{-1}(\bar{q}(k+1))} F(x)\,\mathrm{d}x\right)\right]
\end{aligned}$$

$$\leq \bar{F}\left[-c+\alpha\left(\underline{\beta}(k)r+(1-\underline{\beta}(k))\int_{-\infty}^{\bar{F}^{-1}(\bar{q}(k))} F(x)\,\mathrm{d}x\right)\right]$$
$$= \bar{F}\left[-c+\alpha\left(\underline{\beta}(k)r+(1-\underline{\beta}(k))\underline{J}(k)\right)\right] = \bar{q}(k).$$

The first inequality follows from $z_1(k; w, \beta, q) \geq \underline{\beta}(k)$ and the assumption that $z_2(k+1; w, \beta, q) \leq \bar{q}(k+1)$ and Lemma 11 that the integral is less than $r$. The second inequality follows from Lemma 21 that $\bar{q}(k+1) < \bar{q}(k)$ and the last two equalities follow from property (iii) of Lemma 21 and (100). Since $z_2(k; w, \beta, q) \geq q_\infty$ by construction, $z_2(k; w, \beta, q) \in \mathscr{Z}_2(k)$.

The following lemma shows the properties of the partial derivatives of $h(\cdot)$ for values in the set $\mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1)$.

**Lemma 23** *For any $\epsilon > 0$, there exist $w_0, M \geq 0$ such that the following holds for $w \geq w_0$ and $(\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1)$:*

$$\frac{\partial h_2(z)}{\partial z_2} \leq 1,\ \ \frac{\partial h_3(z)}{\partial z_2} \leq \frac{\epsilon}{2} q_\infty,\ \ \frac{\partial h_1(z)}{\partial z_1} \leq 1 \ \ \textit{and} \ \ -\frac{\partial h_2(z)}{\partial z_1} \leq M, \tag{103}$$

*where $z = z(w+1; w+1, \beta, q) = (\beta, q, f_{w+1}(\beta, q))$.*

*Proof* We show that the four inequalities hold for $w$ large enough one by one.

We first show that $\partial h_2/\partial z_2 \leq 1$ for $w$ large. Recall from Eq. (77) that for $z \in \mathscr{Z}$,

$$\frac{\partial h_2(z)}{\partial z_2} = \alpha(1-\beta)(1-q)\frac{f[\bar{F}^{-1}(h_2(z))]}{f(\bar{F}^{-1}(q))} \leq \alpha(1-q_\infty)\frac{f[\bar{F}^{-1}(h_2(z))]}{f(\bar{F}^{-1}(q))}, \tag{104}$$

where the inequality follows because $q \geq q_\infty$ and $\beta \in \mathscr{Z}_1(w+1) \subseteq (0, 1]$.

By continuity of $f(\cdot)$ and $\bar{F}^{-1}(\cdot)$ at $q_\infty$, there exists $\delta_1 > 0$ such that for all $x$ such that $|x - q_\infty| < \delta_1$, the following holds:

$$f(\bar{F}^{-1}(q_\infty))\sqrt{1-q_\infty} \leq f(\bar{F}^{-1}(x)) \leq \frac{1}{\sqrt{1-q_\infty}} f(\bar{F}^{-1}(q_\infty)). \tag{105}$$

It follows from Lemma 21 that $\bar{q}(w) \to q_\infty$ as $w \to \infty$. Thus, there exists $w_1 \geq 1$ such that $|\bar{q}(w) - q_\infty| < \delta_1$ for $w \geq w_1$. In particular,

$$|\bar{q}(w+1) - q_\infty| < \delta_1 \ \text{ and } \ |\bar{q}(w) - q_\infty| < \delta_1,\ \ w \geq w_1. \tag{106}$$

It follows from (83) that $h_2(z) = z_2(w; w+1, \beta, q)$. We have that $h_2(z) \in \mathscr{Z}_2(w) = [q_\infty, \bar{q}(w)]$ by Lemma 22. In addition, by assumption, $q \in \mathscr{Z}_2(w+1) = [q_\infty, \bar{q}(w+1)]$. Thus, it follows from (106) that for $w \geq w_1$ and $(\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1)$,

$$|q - q_\infty| < \delta_1 \ \text{ and } \ |h_2(z) - q_\infty| < \delta_1.$$

By Eq. (105), we have that for $w \geq w_1$ and $(\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1)$,

$$\frac{f[\bar{F}^{-1}(h_2(z))]}{f(\bar{F}^{-1}(q))} \leq \frac{1}{1-q_\infty}.$$

Substituting this inequality into Eq. (104), we obtain the following:

$$\frac{\partial h_2(z)}{\partial z_2} \leq \alpha \leq 1 \ \text{ for } \ w \geq w_1 \ \text{ and } \ (\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1),$$

where $z = (\beta, q, f_{w+1}(\beta, q)) = z(w+1; w+1, \beta, q)$.

Next we show that $\partial h_3 / \partial z_2 \leq \epsilon q_\infty / 2$ for $w$ large. It follows from Eq. (80) that

$$\frac{\partial h_3(z)}{\partial z_2} = \frac{z_3(w+1; w+1, \beta, q)}{(1-q)^2} \leq \frac{(1-q_\infty)^{w+1}}{(1-q)^2} \leq \frac{(1-q_\infty)^{w}}{(1-\bar{q}(w_1))^2}, \tag{107}$$

where the first inequality follows from Lemma 16. Recall from Lemma 21 that $\bar{q}(w)$ is decreasing. Thus, the second inequality holds because

$$q \leq \bar{q}(w+1) \leq \bar{q}(w_1), \ w \geq w_1.$$

Since $1 - q_\infty < 1$, there exists a constant $w_2 \geq w_1$ such that

$$(1-q_\infty)^w \leq \frac{\epsilon}{2} q_\infty (1-\bar{q}(w_1))^2, \ w \geq w_2.$$

Substituting this inequality into (107), we have the following:

$$\frac{\partial h_3(z)}{\partial z_2} \leq \frac{\epsilon}{2} q_\infty \ \text{ for } \ w \geq w_2 \ \text{ and } \ (\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1).$$

We then show that $\partial h_1 / z_1 \leq 1$ for $w$ large enough. Recall from Eq. (73) that for $w \geq w_2$,

$$0 \leq \frac{\partial h_1(z)}{\partial z_1} = \frac{h_1^2(z)}{\beta^2} \frac{1-b}{1 - a z_3(w+1; w+1, \beta, q)} \leq \frac{h_1^2(z)}{\beta^2} \frac{1-b}{1-a(1-q_\infty)^{w+1}}, \tag{108}$$

where the inequality follows from Lemma 16. Note by assumption that $\beta \in \mathscr{Z}_1(w+1) = [\underline{\beta}(w+1), b]$ and Lemma 22 that $h_1(z) = z_1(w; w+1, \beta, q) \in \mathscr{Z}_1(w) = [\underline{\beta}(w), b]$. In other words, the following holds:

$$\underline{\beta}(w+1) \leq \beta \leq b \ \text{ and } \ \underline{\beta}(w) \leq h_1(z) \leq b.$$

It follows from Lemma 21 that $\underline{\beta}(w) \to b$ as $w \to \infty$. In addition, $1 - a(1 - q_\infty)^{w+1} \to 1$ as $w \to \infty$. Thus, there exists $w_3 \geq w_2$ such that

$$\frac{b}{\underline{\beta}(w+1)} \leq \frac{1}{\sqrt[4]{1-b}} \ \text{ and } \ 1 - a(1-q_\infty)^{w+1} > \sqrt[4]{1-b}, \ w \geq w_3.$$

Substituting these two inequalities into (108), we have that for $w \geq w_3$ and $(\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1)$,

$$0 \leq \frac{\partial h_1(z)}{\partial z_1} \leq \frac{h_1^2(z)}{\beta^2} \frac{1-b}{1-a(1-q_\infty)^{w+1}}$$
$$\leq \frac{b^2}{\underline{\beta}^2(w+1)} \frac{1-b}{1-a(1-q_\infty)^{w+1}} \leq \sqrt[4]{1-b} < 1.$$

Lastly, we show that there exists $M \geq 0$ such that $\partial h_2 / \partial z_1 \leq M$ for $w$ large enough. It follows from Eq. (76) that for all $w \geq w_3$ and $(\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1)$,

$$\begin{aligned} -\frac{\partial h_2(z)}{\partial z_1} &= f(\bar{F}^{-1}(h_2(z)))\alpha \left( r - \int_{-\infty}^{\bar{F}^{-1}(q)} F(x)\,\mathrm{d}x \right) \frac{\partial h_1(z)}{\partial z_1} \\ &\leq f(\bar{F}^{-1}(h_2(z))\alpha r \frac{\partial h_1(z)}{\partial z_1} \leq f(\bar{F}^{-1}(h_2(z))\alpha r. \end{aligned} \tag{109}$$

The first inequality follows from Lemma 11. The second inequality follows from the first inequality in (103). It follows from the continuity of $f(\cdot)$ and $\bar{F}^{-1}(\cdot)$ that $f(\bar{F}^{-1}(x))$ is bounded on $[q_\infty, \bar{q}(w_3)]$. Recall that $h_2(z) \in \mathscr{Z}_2(w) = [q_\infty, \bar{q}(w)]$. Since $\bar{q}(w)$ is decreasing in $w$ by Lemma 21, it follows that

$$q_\infty \leq h_2(z) \leq \bar{q}(w) \leq \bar{q}(w_3), \ \ w \geq w_3 \ \text{ and } \ (\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1).$$

Thus, the right-hand side of the third line in Eq. (109) is bounded. Letting $M$ denote one such bound completes the proof.

In summary, letting $w_0 = w_3$, the four inequalities in (103) hold for all $w \geq w_0$ and $(\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1)$

The next lemma is key to proving Lemma 5.

**Lemma 24** *The following holds:*

$$\lim_{w\to\infty} \sup \left\{ \left| \frac{\partial f_w(\beta, q)}{\partial \beta} \right| : (\beta, q) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w) \right\} = 0, \tag{110}$$

$$\lim_{w\to\infty} \sup \left\{ \left| \frac{\partial f_w(\beta, q)}{\partial q} \right| : (\beta, q) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w) \right\} = 0. \tag{111}$$

*Proof* We first show (110). To facilitate the analysis to follow, define a sequence $y_w$ for $w \geq 1$ as follows:

$$y_w = \sup \left\{ -\frac{\partial f_w(\beta, q)}{\partial q} : (\beta, q) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w) \right\}. \tag{112}$$

It follows from Eq. (94) that $y_w \geq 0$ for all $w \geq 1$. In addition, it follows from Eq. (90) that for any $(\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1)$ and $w \geq 1$,

$$
\begin{aligned}
-\frac{\partial f_{w+1}(\beta, q)}{\partial q} &= \frac{\dfrac{\partial f_w(e(w))}{\partial q}\dfrac{\partial h_2(z(w+1))}{\partial z_2} - \dfrac{\partial h_3(z(w+1))}{\partial z_2}}{\dfrac{\partial f_w(e(w))}{\partial \beta}\dfrac{\partial h_1(z(w+1))}{\partial z_3} + \dfrac{\partial f_w(e(w))}{\partial q}\dfrac{\partial h_2(z(w+1))}{\partial z_3} - \dfrac{\partial h_3(z(w+1))}{\partial z_3}} \\
&= \frac{-\dfrac{\partial f_w(e(w))}{\partial q}\dfrac{\partial h_2(z(w+1))}{\partial z_2} + \dfrac{\partial h_3(z(w+1))}{\partial z_2}}{-\dfrac{\partial f_w(e(w))}{\partial \beta}\dfrac{\partial h_1(z(w+1))}{\partial z_3} - \dfrac{\partial f_w(e(w))}{\partial q}\dfrac{\partial h_2(z(w+1))}{\partial z_3} + \dfrac{\partial h_3(z(w+1))}{\partial z_3}} \\
&\leq \left(-\frac{\partial f_w(e(w))}{\partial q}\frac{\partial h_2(z(w+1))}{\partial z_2} + \frac{\partial h_3(z(w+1))}{\partial z_2}\right) \Big/ \frac{\partial h_3(z(w+1))}{\partial z_3} \\
&\leq \left(y_w \frac{\partial h_2(z(w+1))}{\partial z_2} + \frac{\partial h_3(z(w+1))}{\partial z_2}\right) \Big/ \frac{\partial h_3(z(w+1))}{\partial z_3},
\end{aligned}
\tag{113}
$$

where $z(k) = z(k; w+1, \beta, q)$ for $k = w, w+1$ and $e(w) = (z_1(w), z_2(w))$. We flip the signs of the terms in the second line of the right-hand side. Thus, it follows from Lemma 12 that every term in both the numerator and the denominator (of the right-hand side of the third line) is positive. This leads to the inequality in the fourth line. The last inequality follows from Eq. (112) because $(z_1(w), z_2(w)) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$, which in turn follows from Lemma 22.

Rewriting Eq. (113) gives

$$
-\frac{\partial f_{w+1}(\beta, q)}{\partial q} \leq \left(y_w \frac{\partial h_2(z(w+1))}{\partial z_2} + \frac{\partial h_3(z(w+1))}{\partial z_2}\right) \Big/ \frac{\partial h_3(z(w+1))}{\partial z_3}.
$$

Taking the supremum of both sides over $(\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1)$ gives the following:

$$
\begin{aligned}
y_{w+1} &\leq \sup_{(\beta,q)\in\mathscr{Z}_1(w+1)\times\mathscr{Z}_2(w+1)} \left[\left(y_w \frac{\partial h_2(z(w+1))}{\partial z_2} + \frac{\partial h_3(z(w+1))}{\partial z_2}\right) \Big/ \frac{\partial h_3(z(w+1))}{\partial z_3}\right] \\
&\leq \sup_{(\beta,q)\in\mathscr{Z}_1(w+1)\times\mathscr{Z}_2(w+1)} (1 - q_\infty) \left(y_w \frac{\partial h_2(z(w+1))}{\partial z_2} + \frac{\partial h_3(z(w+1))}{\partial z_2}\right),
\end{aligned}
\tag{114}
$$

where the last inequality follows from Eq. (81) and that $q \in [q_\infty, \bar{q}(w+1)]$. In particular,

$$
\frac{\partial h_3(z(w+1))}{\partial z_3} = \frac{1}{1 - z_2(w+1)} = \frac{1}{1-q} \geq \frac{1}{1-q_\infty}. \tag{115}
$$

Substituting the first two inequalities in Eqs. (103) into (114) yields the following:

$$
y_{w+1} \leq (1 - q_\infty)\left(y_w + \frac{\epsilon}{2} q_\infty\right) \quad \text{for} \quad w \geq w_0,
$$

where $w_0$ is as in Lemma 23. By induction, we obtain the following inequality: For all $w \geq w_0$ and $n \geq 1$,

$$
\begin{aligned}
y_{w+n} \leq & (1-q_\infty) y_{w+n-1} + \frac{\epsilon}{2} q_\infty (1-q_\infty) \\
\leq & (1-q_\infty)^2 y_{w+n-2} + \frac{\epsilon}{2} q_\infty (1-q_\infty)^2 + \frac{\epsilon}{2} q_\infty (1-q_\infty) \\
\leq & \cdots \leq (1-q_\infty)^n y_w + \frac{\epsilon}{2} q_\infty \sum_{i=1}^{n} (1-q_\infty)^i \\
= & (1-q_\infty)^n y_w + \frac{\epsilon}{2}.
\end{aligned}
$$

Now fix $w = w_0$. There exists $n_1$ such that for all $n \geq n_1$, $(1-q_\infty)^n y_{w_0} < \epsilon/2$. That is, for any $w \geq w_0 + n_1$, $y_w < \epsilon$. Therefore, $y_w \to 0$, as $w \to \infty$. Since $y_w \geq 1$ for all $w \geq 1$, we deduce Eq. (110).

Next, we prove (111) in a similar fashion. Define a sequence $x_w$ as follows:

$$
x_w = \sup \left\{ \frac{\partial f_w(\beta, q)}{\partial \beta} : (\beta, q) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w) \right\}, \ w \geq 1.
$$

It follows from Eq. (94) that $x_w \geq 0$ for all $w \geq 1$. In addition, it follows from Eq. (89) that the following holds: For any $w \geq 1$ and $(\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1)$

$$
\begin{aligned}
\frac{\partial f_{w+1}(\beta, q)}{\partial \beta} &= - \frac{\frac{\partial f_w(e(w))}{\partial \beta} \frac{\partial h_1(z(w+1))}{\partial z_1} + \frac{\partial f_w(e(w))}{\partial q} \frac{\partial h_2(z(w+1))}{\partial z_1}}{\frac{\partial f_w(e(w))}{\partial \beta} \frac{\partial h_1(z(w+1))}{\partial z_3} + \frac{\partial f_w(e(w))}{\partial q} \frac{\partial h_2(z(w+1))}{\partial z_3} - \frac{\partial h_3(z(w+1))}{\partial z_3}} \\
&= \frac{\frac{\partial f_w(e(w))}{\partial \beta} \frac{\partial h_1(z(w+1))}{\partial z_1} + \frac{\partial f_w(e(w))}{\partial q} \frac{\partial h_2(z(w+1))}{\partial z_1}}{-\frac{\partial f_w(e(w))}{\partial \beta} \frac{\partial h_1(z(w+1))}{\partial z_3} - \frac{\partial f_w(e(w))}{\partial q} \frac{\partial h_2(z(w+1))}{\partial z_3} + \frac{\partial h_3(z(w+1))}{\partial z_3}} \\
&\leq \left( \frac{\partial f_w(e(w))}{\partial \beta} \frac{\partial h_1(z(w+1))}{\partial z_1} + \frac{\partial f_w(e(w))}{\partial q} \frac{\partial h_2(z(w+1))}{\partial z_1} \right) \Big/ \frac{\partial h_3(z(w+1))}{\partial z_3} \\
&\leq \left( x_w \frac{\partial h_1(z(w+1))}{\partial z_1} + y_w \left( - \frac{\partial h_2(z(w+1))}{\partial z_1} \right) \right) \Big/ \frac{\partial h_3(z(w+1))}{\partial z_3}.
\end{aligned}
$$

The first inequality holds because every term in both the numerator and denominator of the right-hand side of the second line is positive. The last inequality holds because $(z_1(w), z_2(w)) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$. Taking the supremum of both sides over $(\beta, q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1)$ yields the following: For $w \geq 1$,

$$
x_{w+1} \leq \sup_{(\beta,q) \in \mathscr{Z}_1(w+1) \times \mathscr{Z}_2(w+1)} \left( x_w \frac{\partial h_1(z(w+1))}{\partial z_1} - y_w \frac{\partial h_2(z(w+1))}{\partial z_1} \right) \Big/ \frac{\partial h_3(z(w+1))}{\partial z_3}.
$$

Substituting the last two inequalities in Eqs. (103) and (115) into this inequality yields the following:

$$
x_{w+1} \leq (1-q_\infty)(x_w + M y_w) \ \text{ for } \ w \geq w_0.
$$

We have just shown that $y_w \to 0$ as $w \to \infty$. Therefore, for $\epsilon > 0$, there exists $w_1 \geq w_0$ such that $y_w < q_\infty \epsilon / 2M$ for $w \geq w_1$. Thus, the following holds:

$$x_{w+1} \leq (1 - q_\infty) \left( x_w + \frac{\epsilon}{2} q_\infty \right) \text{ for } w \geq w_1.$$

Fixing $w = w_1$, we have that

$$\begin{aligned} x_{w_1+n} &\leq (1 - q_\infty) x_{w_1+n-1} + \frac{\epsilon}{2} q_\infty (1 - q_\infty) \\ &\leq (1 - q_\infty)^2 x_{w_1+n-2} + \frac{\epsilon}{2} q_\infty (1 - q_\infty)^2 + \frac{\epsilon}{2} q_\infty (1 - q_\infty) \\ &\leq \cdots \leq (1 - q_\infty)^n x_{w_1} + \frac{\epsilon}{2} q_\infty \sum_{i=1}^{n} (1 - q_\infty)^i \\ &\leq (1 - q_\infty)^n x_{w_1} + \frac{\epsilon}{2}. \end{aligned}$$

There exists $n_2$ such that for $n \geq n_2$, $(1 - q_\infty)^n x_{w_1} < \epsilon/2$. Thus, for $w \geq w_1 + n_2$, $x_w < \epsilon$. Since $x_w \geq 1$, we have that $\lim_{w \to \infty} x_w = 0$, which gives (111).

### Characterizing the equilibrium quantities with $f_w(\cdot)$

This subsection relates the equilibrium quantities to $f_w(\cdot)$. The following lemma shows that the equilibrium $e^*$ can be characterized by $h(\cdot)$.

**Lemma 25** *For any equilibrium $e^*$, the following holds:*

$$(\beta^*(w), q^*(w), \bar{G}^*(w)) = h(\beta^*(w+1), q^*(w+1), \bar{G}^*(w+1)), \ w \geq 1.$$

*Proof* It follows from (4) to (70) that $\beta^*(w) = h_1(\beta^*(w+1), q^*(w+1), \bar{G}^*(w+1))$ for $w \geq 1$.

It follows from Eq. (11) that

$$J^*(w) = \mathbb{E}_\varepsilon [\bar{F}^{-1}(q^*(w)) - (\varepsilon(1) - \varepsilon(0))]^+ = \int_{-\infty}^{\bar{F}^{-1}(q^*(w))} F(x) \, \mathrm{d}x.$$

Substituting this equation and $\beta^*(w) = h_1(\beta^*(w+1), q^*(w+1), \bar{G}^*(w+1))$ into (10) and comparing it with (71), we have that

$$\begin{aligned} q^*(w) &= \bar{F} \left( -c + \alpha \left\{ \beta^*(w) r + (1 - \beta^*(w)) \int_{-\infty}^{\bar{F}^{-1}(q^*(w))} F(x) \, \mathrm{d}x \right\} \right) \\ &= h_2(\beta^*(w+1), q^*(w+1), \bar{G}^*(w+1)). \end{aligned}$$

In addition, it follows from Eqs. (1) to (72) that

$$\begin{aligned}\bar{G}^*(w) &= \frac{\bar{G}^*(w+1)}{1-q^*(w+1)} = \min\left(1, \frac{\bar{G}^*(w+1)}{1-q^*(w+1)}\right) \\ &= h_3(\beta^*(w+1), q^*(w+1), \bar{G}^*(w+1)),\end{aligned}$$

where the second equality holds because $\bar{G}^*(w) = \prod_{i=1}^{w}(1-q^*(i)) \leq 1$.

Thus, it is immediate that $f_w(\cdot)$ characterizes $\bar{G}^*(w)$ in terms of $\beta^*(w)$ and $q^*(w)$, which is formalized in the following corollary.

**Corollary 5** *We have that* $\bar{G}^*(w) = f_w(e^*(w))$ *for all* $w \geq 1$.

*Proof* Fixing a $w \geq 1$ and substituting $z(w) = (\beta^*(w), q^*(w), \bar{G}^*(w))$ into Eq. (83), by applying Lemma 25 inductively we have that $z(1) = (\beta^*(1), q^*(1), \bar{G}^*(1))$. Note that the resulting $z(1) = (\beta^*(1), q^*(1), \bar{G}^*(1))$ satisfies condition (84). In particular, $\bar{G}^*(1) = 1 - q^*(1)$. Thus, it follows from the definition of the function $f_w(\cdot)$ that $\bar{G}^*(w) = f_w(e^*(w))$ for all $w \geq 1$.

The following lemma shows that the equilibrium quantities live in the set $\mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$.

**Lemma 26** *For any equilibrium* $e^* = (\beta^*, q^*)$*, we have that*

$$e^*(w) = (\beta^*(w), q^*(w)) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w), \; w \geq 1.$$

*Proof* By Lemma 4 and Corollaries 3–4, $\beta^*(w) \leq b$ and $q^*(w) \geq q_\infty$. Thus, it suffices to show that $\beta^*(w) \geq \underline{\beta}(w)$ and $q^*(w) \leq \bar{q}(w)$ for all $w \geq 1$.

We first show that $\beta^*(w) \geq \underline{\beta}(w)$. It follows from Proposition 1 that

$$\begin{aligned}\beta^*(w) &= \left(1 + \sum_{t=w}^{\infty}\prod_{i=w}^{t}\frac{1-b}{1-a\bar{G}^*(i+1)}\right)^{-1} \\ &\geq \left(1 + \sum_{t=w}^{\infty}\prod_{i=w}^{t}\frac{1-b}{1-a\bar{G}^*(w)}\right)^{-1} \\ &= \left(1 + \sum_{i=1}^{\infty}\left(\frac{1-b}{1-a\bar{G}^*(w)}\right)^{i}\right)^{-1} \\ &= 1 - \frac{1-b}{1-a\bar{G}^*(w)} \\ &\geq 1 - \frac{1-b}{1-a(1-q_\infty)^w} = \frac{b-a(1-q_\infty)^w}{1-a(1-q_\infty)^w} = \underline{\beta}(w),\end{aligned}$$

where the first inequality follows from

$$\bar{G}^*(i+1) = \bar{G}^*(w)\prod_{j=w+1}^{i+1}(1-q^*(j)) \leq \bar{G}^*(w), \; i \geq w,$$

and the last inequality follows from Lemma 16. In particular, it follows from

$$\bar{G}^*(w) = z_3(w; w, \beta^*(w), q^*(w)) \le (1 - q_\infty)^w.$$

Therefore, we have that $\beta^*(w) \in \mathscr{Z}_1(w)$.

We then prove that $q^*(w) \in \mathscr{Z}_2(w)$. We first show that $J^*(w) \ge \underline{J}(w)$ for all $w \ge 1$ by contradiction, where $J^*(w)$ is the expected discounted utility of waiting. Suppose this is not true and there exists $w_0$ such that $J^*(w_0) < \underline{J}(w_0)$.

We first show by induction that $J^*(w) < \underline{J}(w_0)$ for all $w \ge w_0$. It is true for $w_0$ by assumption. By the induction hypothesis, suppose it is true for $k = w$. In particular, $J^*(w) < \underline{J}(w_0)$. We then show that $J^*(w+1) < \underline{J}(w_0)$. Substituting Eq. (8) and $\underline{J}(w_0) = \kappa(\underline{\beta}(w_0), \underline{J}(w_0))$ into $J^*(w) < \underline{J}(w_0)$, we obtain that

$$\begin{aligned}
&\mathbb{E}[-c + \alpha[\beta^*(w)r + (1-\beta^*(w))J^*(w+1)] - (\varepsilon(1) - \varepsilon(0))]^+ \\
&\quad = \mathbb{E}[\max\{\varepsilon(1), -c + \alpha[\beta^*(w)r + (1-\beta^*(w))J^*(w+1)] + \varepsilon(0)\}] \\
&\quad = J^*(w) < \underline{J}(w_0) \\
&\quad = \mathbb{E}[-c + \alpha[\underline{\beta}(w_0)r + (1-\underline{\beta}(w_0))\underline{J}(w_0)] - (\varepsilon(1) - \varepsilon(0))]^+,
\end{aligned}$$

where the first equality holds because $\mathbb{E}[\varepsilon(1)] = 0$. Comparing the first line and right-hand side of the last line, we conclude that the following inequality holds:

$$\begin{aligned}
&\beta^*(w)r + (1-\beta^*(w))J^*(w+1) \\
&\quad < \underline{\beta}(w_0)r + (1-\underline{\beta}(w_0))\underline{J}(w_0) \\
&\quad = \beta^*(w)r + (1-\beta^*(w))\underline{J}(w_0) + (\underline{\beta}(w_0) - \beta^*(w))(r - \underline{J}(w_0)).
\end{aligned}$$

Rearranging the terms, we have that

$$J^*(w+1) < \underline{J}(w_0) + \frac{(\underline{\beta}(w_0) - \beta^*(w))(r - \underline{J}(w_0))}{1 - \beta^*(w)}. \tag{116}$$

Note that the last term on the right-hand side of Eq. (116) is non-positive. To see this, recall that we have shown $\beta^*(w) \ge \underline{\beta}(w)$ at the beginning of this proof. It follows from property (i) of Lemma 21 that $\underline{\beta}(w) \ge \underline{\beta}(w_0)$ for $w \ge w_0$. Combining the two inequalities, we have that $\underline{\beta}(w_0) - \beta^*(w) \le 0$. In addition, it follows from Lemma 21 that $r - \underline{J}(w_0) \ge 0$. Since $1 - \hat{\beta}^*(w) > 0$, we have that the last term on the right-hand side of Eq. (116) is non-positive.

By dropping the non-positive term on the right-hand side of (116), we have that $J^*(w+1) < \underline{J}(w_0)$, which completes the induction argument. In summary, $J^*(w) < \underline{J}(w_0)$ for all $w \ge w_0$.

On the one hand, by the induction argument, we prove that $J^*(w) < \underline{J}(w_0)$ for all $w \ge w_0$. Thus, it follows from Corollary 4 that $J_\infty = \lim_{w\to\infty} J^*(w) \le \underline{J}(w_0)$. On the other hand, it follows from properties (i)–(ii) in Lemma 21 that $\underline{J}(w_0) < J_\infty$. This leads to a contradiction. Therefore, there exists no $w_0$ such that $J^*(w_0) < \underline{J}(w_0)$. In other words, $J^*(w) \ge \underline{J}(w)$ for all $w \ge 1$.

We complete the proof by showing $q^*(w) \leq \bar{q}(w)$ for $w \geq 1$. It follows from Eq. (11) that

$$J^*(w) = \mathbb{E}_\varepsilon[\bar{F}^{-1}(q^*(w)) - (\varepsilon(1) - \varepsilon(0))]^+ = \int_{-\infty}^{\bar{F}^{-1}(q^*(w))} F(x)\,\mathrm{d}x.$$

In addition, recall from Lemma 21 that

$$\underline{J}(w) = \int_{-\infty}^{\bar{F}^{-1}(\bar{q}(w))} F(x)\,\mathrm{d}x.$$

By comparing these two equations, we can conclude that $q^*(w) \leq \bar{q}(w)$ because $J^*(w) \geq \underline{J}(w)$ for $w \geq 1$. □

**Proof of Lemma 5**

It follows from Corollary 5 that

$$\bar{G}_i^*(w) = f_w(\beta_i^*(w), q_i^*(w)),\ i = 1, 2 \ \text{ and } \ w \geq 1.$$

Applying the mean value theorem for multivariable functions to $f_w(\cdot)$, the following holds: For $w \geq 1$,

$$\begin{aligned} \delta_{\bar{G}}(w) =& f_w((\beta_1^*(w), q_1^*(w))) - f_w((\beta_2^*(w), q_2^*(w))) \\ =& \frac{\partial f_w(\tilde{e}(w))}{\partial q}\delta_q(w) + \frac{\partial f_w(\tilde{e}(w))}{\partial \beta}\delta_\beta(w), \end{aligned} \tag{117}$$

where $\tilde{e}(w) = C(w)e_1^*(w) + (1 - C(w))e_2^*(w)$ for some $C(w) \in (0, 1)$. It follows from Lemma 26 that $e_1^*(w), e_2^*(w) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$. Since $\mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$ is convex, $\tilde{e}(w) \in \mathscr{Z}_1(w) \times \mathscr{Z}_2(w)$ for all $w \geq 1$. It follows from Lemma 24 that

$$\lim_{w\to\infty} \frac{\partial f_w(\tilde{e}(w))}{\partial q} = 0 \ \text{ and } \ \lim_{w\to\infty} \frac{\partial f_w(\tilde{e}(w))}{\partial \beta} = 0.$$

Thus, we conclude that for any $\epsilon > 0$, there exists a nonnegative constant $w_1$ such that the following inequalities are satisfied:

$$\left|\frac{\partial f_w(\tilde{e}(w))}{\partial q}\right| \leq \epsilon \ \text{ and } \ \left|\frac{\partial f_w(\tilde{e}(w))}{\partial \beta}\right| \leq \epsilon,\ w \geq w_1.$$

Substituting the two inequalities into Eq. (117), we obtain that

$$|\delta_{\bar{G}}(w)| \leq \epsilon(|\delta_\beta(w)| + |\delta_q(w)|),\ w \geq w_1.$$

□

## Appendix 3: The road map for the proof of uniqueness

This appendix provides a detailed road map of the uniqueness proof (Proposition 4). The proof is done by contradiction. In what follows, we first provide an overview of the key steps that lead to the contradiction using various auxiliary lemmas (see Fig. 8). We then summarize the key steps to proving Lemma 5 provided in Appendix 2, which is an important technical lemma for the uniqueness proof. Two auxiliary functions, denoted by $h(\cdot)$ and $f_w(\cdot)$, and several lemmas in Appendix 2 facilitate the proof of Lemma 5. Figure 9 provides a diagram to show how the lemmas in Appendix 2 are used to prove Lemma 5.

The proof (of uniqueness) by contradiction proceeds as follows: Suppose that there are two different equilibria and define their difference as $(\delta_\beta, \delta_q)$. The contradiction is built on the limiting properties of the difference $(\delta_\beta(w), \delta_q(w))$ as $w$ tends to infinity. Figure 8 shows how the contradiction is constructed.

On the one hand, Corollaries 3–4 provide the limits of equilibrium quantities (in any potential equilibrium). It is immediate from these two corollaries that the difference of the equilibrium quantities (of two different equilibria) vanishes as $w$ goes to infinity, i.e.,

$$(\delta_\beta(w), \delta_q(w)) \to 0 \;\text{ as }\; w \to \infty.$$

On the other hand, Lemmas 7 and 8 show that this convergence cannot hold. Lemma 7 shows that the difference of the equilibrium quantities $(\delta_\beta(w), \delta_q(w))$, $w \geq 0$, is characterized by a dynamical system. To be more specific, the function characterizing the evolution of this dynamical system has two parts: a constant matrix $A$ with a special structure and a matrix function $B(\cdot)$. In addition, the perturbation function $B(w)$ vanishes as $w$ goes to infinity, i.e., $|||B(w)|||_\infty \to 0$ as $w \to \infty$. This property is proved with the help of the technical Lemma 5. Then Lemma 8 shows that the dynamical system given in Lemma 7 cannot converge to zero. Combining Lemmas 7 and 8, we conclude that the difference $(\delta_\beta(w), \delta_q(w)))$ cannot converge to zero, which leads to the contradiction.

The rest of this section summarizes the critical steps in Appendix 2 to prove the technical Lemma 5, which characterizes $\delta_{\bar{G}}(w)$ in terms of $\delta_\beta(w)$ and $\delta_q(w)$ for $w \geq 0$ using the functions $g_1(\cdot)$ and $g_2(\cdot)$. Figure 9 illustrates how various lemmas are used (and relate to one another) to prove Lemma 5. To be specific, "Definition of the auxiliary function $f_w(\cdot)$," "Partial derivatives of the auxiliary function $f_w(\cdot)$" and "Properties of $f_w(\cdot)$ on a restricted set" in Appendix 2 construct two auxiliary functions $h(\cdot)$ and $f_w(\cdot)$ and provide various properties of these two functions. "Characterizing the equilibrium quantities with $f_w(\cdot)$" in Appendix 2 shows the characterization of the equilibrium quantities using the auxiliary functions $h(\cdot)$ and $f_w(\cdot)$. Thus, the properties of the auxiliary functions provided in "Definition of the auxiliary function $f_w(\cdot)$," "Partial derivatives of the auxiliary function $f_w(\cdot)$," and "Properties of $f_w(\cdot)$ on a restricted set" in Appendix 2 are applicable to the equilibrium quantities. Appendix 2 proves Lemma 5.

The proof of Lemma 5 in Appendix 2 includes two parts. The first part constructs the functions $g_1(\cdot)$ and $g_2(\cdot)$ in two steps. In the first step, we use the auxiliary function $f_w(\cdot)$ defined in "Definition of the auxiliary function $f_w(\cdot)$" in Appendix 2 to

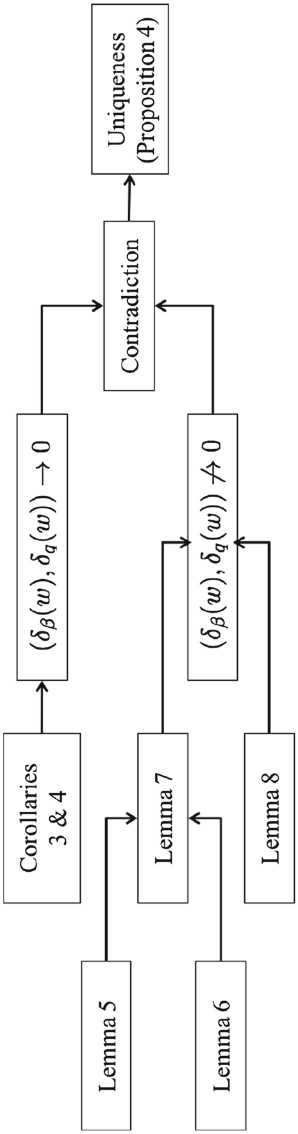


**Fig. 8** The logic flow for proving uniqueness of the equilibrium

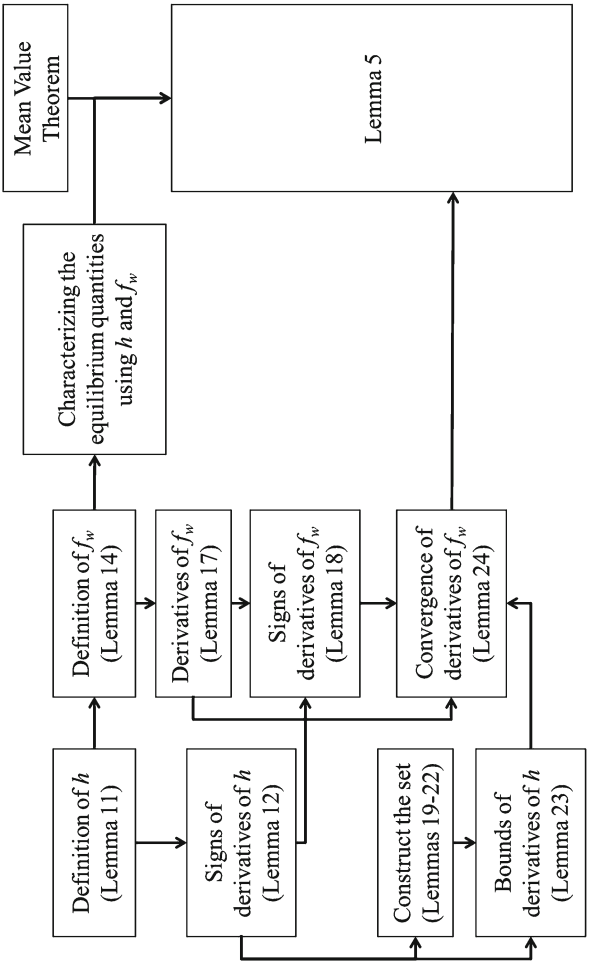


**Fig. 9** The logic flow for proving Lemma 5

characterize $\bar{G}(w)$ in terms of $\beta(w)$ and $q(w)$, i.e., $\bar{G} = f_w(\beta(w), q(w))$ for $w \geq 0$; see Corollary 5 in "Characterizing the equilibrium quantities with $f_w(\cdot)$" in Appendix 2. In the second step, we apply the mean value theorem and construct the functions $g_1$ and $g_2$ using the partial derivatives of the function $f_w(\cdot)$; see Eq. (117).

The second part of the proof of Lemma 5 shows that the two functions $g_1(w)$ and $g_2(w)$ converge to zero as $w \to \infty$. To show this, it is sufficient to show that the supremum norm of the partial derivatives of the function $f_w(\cdot)$ converges to zero as $w \to \infty$; see Eq. (117). However, this statement is not true in general, but is valid if we restrict the arguments of the function $f_w(\cdot)$ to be in the set $\mathscr{L}_1(w) \times \mathscr{L}_2(w)$ defined in "Properties of $f_w(\cdot)$ on a restricted set" in Appendix 2. The convergence result of the partial derivatives of the function $f_w(\cdot)$ restricted in the set $\mathscr{L}_1(w) \times \mathscr{L}_2(w)$ is given by Lemma 24. In addition, Lemma 26 ensures that the equilibrium quantities lie in the $\mathscr{L}_1(w) \times \mathscr{L}_2(w)$. Thus, applying Lemma 24 completes the second part of the proof of Lemma 5.

"Definition of the auxiliary function $f_w(\cdot)$," "Partial derivatives of the auxiliary function $f_w(\cdot)$," and "Properties of $f_w(\cdot)$ on a restricted set" in Appendix 2 are dedicated to proving Lemma 24 using the auxiliary functions $h(\cdot)$ and $f_w(\cdot)$. To be specific, "Definition of the auxiliary function $f_w(\cdot)$" in Appendix 2 defines the auxiliary functions $h(\cdot)$ and $f_w(\cdot)$. "Partial derivatives of the auxiliary function $f_w(\cdot)$" in Appendix 2 provides the recursive equations to characterize the partial derivatives of the function $f_w(\cdot)$ and the signs of the partial derivatives. "Properties of $f_w(\cdot)$ on a restricted set" in Appendix 2 constructs the restricted set $\mathscr{L}_1(w) \times \mathscr{L}_2(w)$ and proves the convergence of the partial derivatives of the function $f_w(\cdot)$ in the restricted set.

The auxiliary function $f_w(\beta, q)$ is defined implicitly through Eqs. (82)–(84), which are rewritten for convenience as follows:

$$\begin{aligned} z(w) &= (\beta, q, f_w(\beta, q)), \\ z(k-1) &= h(z(k)) \ \text{ for } \ k = w, \ldots, 2, \\ z_2(1) &= 1 - z_3(1). \end{aligned}$$

Lemma 14 ensures that the function $f_w(\beta, q)$ is well-defined. In order to make sense of this definition of the implicit function $f_w(\cdot)$, the auxiliary function $h(\cdot)$ needs to be introduced. The function $h(\cdot)$ is constructed such that it characterizes the time-reversed evolution of the equilibrium quantities; see Lemma 25 in "Characterizing the equilibrium quantities with $f_w(\cdot)$" in Appendix 2. This immediately leads to the observation that if we substitute the equilibrium quantities at time $w$ into $z(w)$ in Eq. (82), i.e., $z(w) = (\beta^*(w), q^*(w), \bar{G}^*(w))$, then the values of $z(k)$ (for $k = w-1, \ldots, 1$) in Eq. (83) equal the equilibrium quantities as well, i.e.,

$$z(k) = (\beta^*(k), q^*(k), \bar{G}^*(k)) \ \text{ for } \ k = w-1, \ldots, 1.$$

In addition, the condition in Eq. (84) is automatically satisfied by the definition of $\bar{G}$ in Eq. (1). Therefore, the function $f_w(\cdot)$ is the implicit function that characterizes the equilibrium quantity $\bar{G}^*(w)$ in terms of $\beta^*(w), q^*(w)$, i.e., $\bar{G}^*(w) =$

$f_w(\beta^*(w), q^*(w))$, $w \geq 0$; see Corollary 5 in "Characterizing the equilibrium quantities with $f_w(\cdot)$" in Appendix 2.

Lemma 24 provides the convergence property of the partial derivatives of the implicit function $f_w(\cdot)$. In order to prove this lemma, we first provide a recursive characterization of the partial derivatives of the implicit function $f_w(\cdot)$. Since the function $f_w(\cdot)$ is defined implicitly by using the function $h(\cdot)$ recursively, the partial derivatives of the implicit function $f_w(\cdot)$ are characterized using the partial derivatives of the function $h(\cdot)$ (provided in Lemma 12) recursively; see Lemma 17. By analyzing the partial derivatives of the functions $f_w(\cdot)$ and $h(\cdot)$ (provided in Lemmas 12 and 17), we provide useful properties of the partial derivatives. These properties eventually lead to the convergence property in Lemma 24; see Fig. 9.

We end this section by providing a comment on Lemma 23. Lemma 23 provides critical bounds of the partial derivatives of the function $h(\cdot)$ to prove Lemma 24. However, these bounds only hold after we restrict the arguments of the function $h(\cdot)$ to the set $\mathscr{L}_1(w) \times \mathscr{L}_2(w)$. The set $\mathscr{L}_1(w) \times \mathscr{L}_2(w)$ is carefully constructed to satisfy two conditions. First, the set is narrow enough such that the bounds in Lemma 23 hold. Second, the set is wide enough to ensure that the equilibrium quantities lie in the set; see Lemma 26.

## References


1. Afèche, P., Sarhangian, V.: Rational abandonment from priority queues: equilibrium strategy and pricing implications. Working paper (2017)
2. Aksin, Z., Armony, M., Mehrotra, V.: The modern call-center: a multi-disciplinary perspective on operations management research. Prod. Oper. Manag. **16**, 665–688 (2007)
3. Aksin, Z., Ata, B., Emadi, S., Su, C.: Structural estimation of callers' delay sensitivity in call centers. Manag. Sci. **59**(12), 2727–2746 (2013)
4. Aksin, Z., Ata, B., Emadi, S., Su, C.: Impact of delay announcements in call centers: an empirical approach. Oper. Res. **65**(1), 242–265 (2017)
5. Aliprantis, C.D., Border, K.C.: Infinite Dimensional Analysis: A Hitchhiker's Guide. Springer, New York (2007)
6. Anderson, S.P., de Palma, A., Thisse, J.: Discrete Choice Theory of Product Differentiation. The MIT Press, Cambridge, MA (1992)
7. Apostol, T.M.: Calculus, vol. 2. Wiley, Hoboken (1969)
8. Armony, M., Shimkin, N., Whitt, W.: The impact of delay announcements in many-server queues with abandonment. Oper. Res. **50**(1), 66–81 (2009)
9. Assaf, D., Haviv, M.: Reneging from processor sharing systems and random queues. Math. Oper. Res. **15**(1), 129–138 (1990)
10. Ata, B., Peng, X.: An equilibrium analysis of a multiclass queue with endogenous abandonments in the conventional heavy traffic regime. Working paper (2017)
11. Ata, B., Skaro, A., Tayur, S.: OrganJet: overcoming geographical disparities in access to deceased donor kidneys in the United States. Manag. Sci. (2017, forthcoming). doi:10.1287/mnsc.2016.2487
12. Ata, B., Tongarlak, M.H.: On scheduling a multiclass queue with abandonments under general delay costs. Queueing Syst. **74**(1), 65–104 (2013)
13. Baccelli, F., Hebuterne, G.: On queues with impatient customers. In: Kylstra, F. (ed.) Performance 81, pp. 159–179. North Holland, Amsterdam (1981)
14. Baccelli, F., Boyer, P., Hebuterne, G.: Single-server queues with impatient customers. Adv. Appl. Probab. **16**, 887–905 (1984)
15. Boxma, O., Perry, D., Stadje, W.: The M/G/1+G queue revisited. Queueing Syst. **67**, 207–220 (2011)
16. Bramson, M.: State space collapse with application to heavy traffic limits for multiclass queueing networks. Queueing Syst. **30**, 89–140 (1998)

17. Brandt, A., Brandt, M.: On the $M(n)/M(n)/s$ queue with impatient calls. Perform. Eval. **35**, 1–18 (1999)
18. Dai, J.G., He, S., Tezcan, T.: Many-server diffusion limits for $G/Ph/n + GI$ queues. Ann. Appl. Probab. **20**(5), 1854–1890 (2010)
19. Del Moral, P., Miclo, L.: Self-interacting Markov chains. Stoch. Anal. Appl. **24**, 615–660 (2006)
20. Finch, P.D.: Deterministic customer impatience in the queueing system GI/M/1. Biometrika **47**, 4552 (1960)
21. Gans, N., Koole, G., Mandelbaum, A.: Telephone call centers: tutorial, review and research prospects. Manuf. Serv. Oper. Manag. **5**, 73–141 (2003)
22. Gavish, B., Schweitzer, P.J.: The Markovian queue with bounded waiting time. Manag. Sci. **23**, 1349–1357 (1977)
23. Hassin, R.: On the optimality of first come last served queues. Econometrica **53**(1), 201–202 (1985)
24. Hassin, R., Haviv, M.: Equilibrium strategies for queues with impatient customers. Oper. Res. Lett. **1995**, 41–45 (1995)
25. Hassin, R., Haviv, M.: To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems. Kluwer Academic Publishers, Berlin (2003)
26. Haviv, M., Ritov, Y.: Homogeneous customers renege from invisible queues at random times under deteriorating waiting conditions. Queueing Syst. **38**, 495–508 (2001)
27. Jennings, O., Pender, J.: Comparisons of standard and ticket queues. Queueing Syst. **84**, 145–202 (2016)
28. Kuzu, K., Xu, S.H., Gao, L.: To wait or not to wait: The theory and practice of ticket queues. Working Paper (2017)
29. Lemmens, B., Nussbaum, N.: Nonlinear Perron-Frobenius Theory. Cambridge University Press, Cambridge (2012)
30. Maglaras, C., Yao, J., Zeevi, A.: Observational learning in queues with abandonments. Working paper.(2017)
31. Mandelbaum, A., Momčilović, P.: A model for rational abandonments from invisible queues. Math. Oper. Res. **37**(1), 41–65 (2012)
32. Mandelbaum, A., Shimkin, N.: Queues with many servers and impatient customers. Queueing Syst. **36**, 141–173 (2000)
33. Naor, P.: The regulation of queue size by levying tolls. Econometrica **37**(1), 15–24 (1969)
34. Reed, J., Ward, A.R.: Approximating the GI/GI/1+GI queue with a nonlinear drift diffusion: Hazard rate scaling in heavy traffic. Math. Oper. Res. **33**(3), 606–644 (2008)
35. Rubino, M., Ata, B.: Dynamic control of a make-to-order, parallel-server system with cancellations. Oper. Res. **57**(1), 94–108 (2009)
36. Rudin, W.: Principles of Mathematical Analysis. McGraw-Hill, New York (1976)
37. Shimkin, N., Mandelbaum, A.: Rational abandonment from tele-queues: nonlinear waiting costs with heterogeneous preferences. Queueing Syst. **47**, 117–146 (2004)
38. Stanford, R.E.: Reneging phenomena in single channel queues. Math. Oper. Res. **4**, 162–178 (1979)
39. Stokey, N.L., Lucas, R.E.: Recursive Methods in Economic Dynamics. Harvard University Press, Cambridge (1989)
40. Ward, A.R.: Asymptotic analysis of queueing systems with reneging: a survey of results for FIFO, single class models. Surv. Math. Oper. Res. Manag. Sci. **16**(1), 1–14 (2011)
41. Ward, A.R., Glynn, P.W.: A diffusion approximation for a $GI/GI/1$ queue with balking or reneging. Queueing Syst. **43**, 371–400 (2005)
42. Zeidler, E.: Nonlinear functional analysis and its applications: I: Fixed-point theorems. Springer, New York (1998)
43. Zohar, E., Mandelbaum, A., Shimkin, N.: Adaptive behavior of impatient customers in tele queues: theory and empirical support. Manag. Sci. **48**(4), 566–583 (2002)