Noname manuscript No.
(will be inserted by the editor)

On the $M_t/M_t/K_t + M_t$ queue in heavy traffic

Anatolii A. Puhalskii
University of Colorado Denver and
Institute for Problems in
Information Transmission, Moscow

Abstract The focus of this paper is on the asymptotics of large-time numbers of customers in time-periodic Markovian many-server queues with customer abandonment in heavy traffic. Limit theorems are obtained for the periodic number-of-customers processes under the fluid and diffusion scalings. Other results concern limits for general time-dependent queues and for time-homogeneous queues in steady state.

1 Introduction

Many-server queues with customer abandonment have been the subject of extensive research, the primary motivation coming from modelling call centres, see, e.g., Garnett, Mandelbaum, and Reiman [4], Whitt [17,18], Zeltyn and Mandelbaum [19], and references therein. Those papers testify to the importance of the asymptotics where both the arrival rate and the number of servers tend to infinity, their ratio being maintained, whereas the service and abandonment rates are kept fixed. Most studied is the case of Poisson arrival processes and exponential service and abandonment times where the arrival, service, and abandonment rates, and the number of servers do not vary with time. Fleming, Simon, and Stolyar [3], assuming critical loading, obtain diffusion-scale limit theorems for the stationary number of customers. Garnett, Mandelbaum, and Reiman [4], also for the critical load, derive fluid- and diffusion-scale limits for the number-of-customers and virtual-waiting-time processes, and for the stationary distributions of those processes. Their other results are concerned with limits for the stationary fractions of abandoning customers and of customers who have to wait in the queue, as well as with computing expectations of functions of the waiting time. Similar asymptotics for the overloaded case are obtained in Whitt [17], who assumes a finite waiting room, and Talreja and Whitt [16]. In addition, Whitt [17] provides insight into the case where the number of servers is much greater than the abandonment rate. Talreja and Whitt [16] also give a proof of the virtual-waiting-time-process limit for the critically loaded queue. The Markovian assumptions are relaxed in Zeltyn and Mandelbaum [19] who study steady-state waiting times. A general framework of Markovian stochastic processing systems with time-varying rates is studied by Mandelbaum, Massey, and Reiman [10] who obtain fluid- and diffusion-scale limits for the number-of-customers processes. They do not require certain loading conditions to hold. The application to many-server queues with abandonment is explored in a series of papers by Mandelbaum, Massey, Reiman, and Stolyar who consider time-varying rates, allow the possibility of retrials, and incorporate virtual-waiting-time processes, see, e.g., Mandelbaum, Massey, Reiman, Rider, and Stolyar [9] and references therein.

The purpose of this paper is a study of Markovian many-server queues with customer abandonment in heavy traffic for a time-periodic case where the arrival, abandonment, and service rates, and the number of servers can be modelled as jointly periodic functions of time. Under those hypotheses, the large-time

University of Colorado Denver
and Institute for Problems
in Information Transmission, Moscow
E-mail: anatolii.puhalskii@ucdenver.edu

distributions of the numbers of customers are periodic. The main result of this paper states that the large-time distributions of the properly scaled and normalised numbers of customers converge to the periodic distribution of a limiting diffusion process which arises as a particular case of the results of Mandelbaum, Massey, and Reiman [10]. The convergence of the periodic one-dimensional distributions is further extended to convergence of the periodic processes. The method of proof consists in establishing convergence of the number-of-customers processes and in checking the tightness of the stationary distributions of embedded discrete-time Markov chains. That makes the results of Mandelbaum, Massey, and Reiman [10] essential. Unfortunately, the proofs there contain flaws, as specified in Remark 1 below. Therefore, before embarking on the analysis of the large-time behaviour, I provide a separate proof of the heavy traffic convergence in distribution of the number-of-customers processes in many-server queues with time-varying rates and abandonment. Unlike the proof of Mandelbaum, Massey, and Reiman [10], who invoke the strong approximation techniques, the proof here relies on the martingale theory of weak convergence which seems to be more suitable for this sort of result. An overview of the general approach and the related literature as well as a heavy-traffic analysis of the time-homogeneous many-server queue with abandonment in critical loading can be found in Whitt, Pang, and Talreja [12]. The part dealing with tightness relies on bounds on the first and second moments of the numbers of customers which are uniform over time and may be of interest in their own right. The approach used can be traced back to Liptser and Shiryayev [8, Theorem 8.3.2] and Smorodinskii [15]. Along with the application to the periodic case, I use the convergence of the processes and the moment bounds in order to establish convergence of the stationary number of customers in the time-homogeneous case for all three possible loads: supercritical, critical, and subcritical. On the one hand, this provides a unified treatment of and a different perspective on the results of Fleming, Simon, and Stolyar [3], Garnett, Mandelbaum, and Reiman [4], and Whitt [17] on the limits of the stationary number of customers. On the other hand, not only are the limits for the one-dimensional stationary distributions obtained, but also limits for the stationary versions of the corresponding processes. In addition, it is shown that allowing the abandonment and service rates to depend on the scaling parameter gives rise to extra terms in the limit distributions.

The rest of the paper is organised as follows. In Section 2, the results on the convergence of the number-of-customers processes are stated and proved (Theorem 1 concerns the fluid scaling and Theorem 2 concerns the diffusion scaling). Section 3 is concerned with the periodic case, the main results being presented in Theorem 3 and Theorem 4. In Section 4, the time-homogeneous case is considered, see Theorem 5 and Theorem 6. The moment bounds are relegated to the appendix, see Lemma 6. This paper is an expanded and corrected version of Puhalskii [13].

Notation and conventions. The set of real numbers is denoted by $\mathbb{R}$, the set of nonnegative reals is denoted by $\mathbb{R}_+$, the set of natural numbers is denoted by $\mathbb{N}$, and the set of whole numbers is denoted by $\mathbb{Z}_+$. For real numbers x and y, $x \wedge y = \min(x, y)$, $x \vee y = \max(x, y)$, $x^+ = x \vee 0$, and $\lfloor x \rfloor$ denotes the integer part; $\mathbf{1}_A$ denotes the indicator function of set A. A real-valued function $(f(t),\, t \in \mathbb{R}_+)$ is said to be strongly majorised by a real-valued function $(g(t),\, t \in \mathbb{R}_+)$ if $f(0) \le g(0)$ and the function $(g(t) - f(t),\, t \in \mathbb{R}_+)$ is nondecreasing. With a slight abuse of notation, this relationship is denoted by $f(t) \prec g(t)$. I will say that a function $(f(t),\, t \in \mathbb{R}_+)$ is T-periodic, where $T > 0$, if $f(t + T) = f(t)$ for all t and that a stochastic process $(X(t),\, t \in \mathbb{R}_+)$ is T-periodic if the distributions of $(X(t + T),\, t \in \mathbb{R}_+)$ and of $(X(t),\, t \in \mathbb{R}_+)$ coincide.

The space of rightcontinuous $\mathbb{R}$-valued functions on $\mathbb{R}_+$ with lefthand limits is denoted by $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ and is endowed with Skorohod's J_1-topology and the Borel σ-algebra. For a function $(x_t,\, t \in \mathbb{R}_+)$ from $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$, x_{t-} represents the lefthand limit at t with the convention that $x_{0-} = 0$ and $\Delta x_t = x_t - x_{t-}$. All stochastic processes are assumed to have trajectories from and are considered as random elements of $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$. Convergence in distribution in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ has a standard meaning. The predictable quadratic variation process of a locally square integrable martingale $(M_t,\, t \in \mathbb{R}_+)$ is denoted by $(\langle M \rangle_t,\, t \in \mathbb{R}_+)$. (For more background in weak convergence theory and martingale theory, the reader is referred to Jacod and Shiryaev [7] and Liptser and Shiryayev [8].) The random entities encountered in the article are defined on a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

2 Convergence of the number-of-customers processes

I will consider a sequence of $M_t/M_t/K_t + M_t$ queues indexed by $n \in \mathbb{N}$. The nth queue is fed by a Poisson process of customers of rate λ_t^n at time t. The customers are served by one of the K_t^n servers on a FCFS basis. They may abandon the queue after an exponentially distributed time with parameter θ^n_t at time t. More specifically, conditioned on the arrival time τ, the distribution function of the time until abandonment is given by $1 - \exp(-\int_0^t \theta^n_{s+\tau}\, ds)$, $t \in \mathbb{R}_+$. Similarly, the service times of the customers are exponential with parameter μ^n_t at time t. A customer in service may be relegated to the head of the queue before her service is complete if the server serving the customer becomes unavailable because K^n_t decreases. In that case, the customer starts service from scratch the next time she enters service. (The specific policy used for choosing the server to be removed is inconsequential for the results obtained below.)

The functions λ^n_t, μ^n_t, and θ^n_t are assumed to be $\mathbb{R}_+$-valued locally integrable functions, i.e., $\int_0^t \lambda^n_s\, ds < \infty$, $\int_0^t \mu^n_s\, ds < \infty$, and $\int_0^t \theta^n_s\, ds < \infty$ for all $t \in \mathbb{R}_+$. The functions K^n_t are $\mathbb{R}_+$-valued and Lebesgue measurable. The number of customers present at time 0, the arrival process, the service times, and the abandonment times are mutually independent.

Let A^n_t denote the number of customer arrivals by time t. As mentioned, the process $A^n = (A^n_t,\, t \in \mathbb{R}_+)$ is a Poisson process with time-varying rate λ^n_t. Customer abandonment will be modelled via independent Poisson processes $R^{n,i} = (R^{n,i}_t,\, t \in \mathbb{R}_+)$, $i \in \mathbb{N}$, of rate θ^n_t at time t and customer service will be modelled via independent Poisson processes $B^{n,i} = (B^{n,i}_t,\, t \in \mathbb{R}_+)$, $i \in \mathbb{N}$, of rate μ^n_t at time t. Let Q^n_t represent the number of customers present at time t. The evolution of the customer population is modelled by the following equation

$$Q^n_t = Q^n_0 + A^n_t - \sum_{i=1}^{\infty} \int_0^t \mathbf{1}_{\{Q^n_{s-} \ge K^n_{s-} + i\}}\, dR^{n,i}_s - \sum_{i=1}^{\infty} \int_0^t \mathbf{1}_{\{Q^n_{s-} \wedge K^n_{s-} \ge i\}}\, dB^{n,i}_s\,. \tag{2.1}$$

For an explanation, the third term on the right represents the number of customers who have abandoned the queue by time t and the last term represents the number of service completions by time t. Informally, all customers in service are arranged in order and the ith customer is assigned Poisson process $B^{n,i}$. A jump of $B^{n,i}$ triggers a service completion. Once that occurs, the customers in service are reordered and are assigned possibly different processes $B^{n,i}$ so that there are no gaps in the sequence of the processes $B^{n,i}$ being used. Due to the memoryless property of the exponential distribution, this reassignment does not affect the service time distributions. The indicator function in the fourth term equals one if and only if a jump of $B^{n,i}$ triggers a service completion, so the jump of that term at time t equals $\sum_{i=1}^{Q^n_{t-} \wedge K^n_{t-}} \Delta B^{n,i}_t$ which is the number of the processes $B^{n,i}$ "being used" that jump at time t. (One may want to keep in mind that at most one of these processes jumps at any given time a.s.) The processes $R^{n,i}$ are associated with the abandonment process in a similar fashion. (Equation (2.1) also applies to nonFCFS service disciplines so long as service is performed when there are customers present. Besides, the customers whose service is interrupted due to a lack of servers do not have to be put necessarily at the head of the queue. The purpose of those assumptions is to make the set-up more specific.)

Equation (2.1) has a unique strong solution whose trajectories belong to $\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+)$ which can be shown by applying an iterative argument on the jump times of the Poisson processes. Let me also note that the infinite series, in fact, represent finite sums.

Let processes $M^{n,A} = (M^{n,A}_t,\, t \in \mathbb{R}_+)$, $M^{n,R} = (M^{n,R}_t,\, t \in \mathbb{R}_+)$, and $M^{n,B} = (M^{n,B}_t,\, t \in \mathbb{R}_+)$ be defined by the relations

$$M^{n,A}_t = A^n_t - \int_0^t \lambda^n_s\, ds, \tag{2.2a}$$

$$M^{n,R}_t = \sum_{i=1}^{\infty} \int_0^t \mathbf{1}_{\{Q^n_{s-} \ge K^n_{s-} + i\}}\, dR^{n,i}_s - \int_0^t \theta^n_s (Q^n_s - K^n_s)^+\, ds, \tag{2.2b}$$

$$M^{n,B}_t = \sum_{i=1}^{\infty} \int_0^t \mathbf{1}_{\{Q^n_{s-} \wedge K^n_{s-} \ge i\}}\, dB^{n,i}_s - \int_0^t \mu^n_s\, (Q^n_s \wedge K^n_s)\, ds. \tag{2.2c}$$

By (2.1), (2.2a), (2.2b), and (2.2c),

$$Q_t^n = Q_0^n + \int_0^t \lambda_s^n\, ds - \int_0^t \theta_s^n \big(Q_s^n - K_s^n\big)^+\, ds - \int_0^t \mu_s^n\, (Q_s^n \wedge K_s^n)\, ds + M_t^{n,A} - M_t^{n,R} - M_t^{n,B}\,. \tag{2.3}$$

The following martingale characterisation plays a key role in subsequent developments. Let $\mathcal{F}_t^n$ denote the completion with respect to $\mathbf{P}$ of the σ-algebra generated by the random variables Q_0^n, A_s^n, $B_s^{n,i}$, and $R_s^{n,i}$, where $s \le t$ and $i \in \mathbb{N}$. The associated filtration is denoted by $\mathbf{F}^n$ so that $\mathbf{F}^n = (\mathcal{F}_t^n,\, t \in \mathbb{R}_+)$. It may be worth noting that Q_t^n is $\mathcal{F}_t^n$-measurable.

Lemma 1 *The processes $M^{n,A}$, $M^{n,R}$, and $M^{n,B}$ are $\mathbf{F}^n$-locally square integrable martingales with respective predictable quadratic variation processes*

$$\langle M^{n,A}\rangle_t = \int_0^t \lambda_s^n\, ds, \tag{2.4a}$$

$$\langle M^{n,R}\rangle_t = \int_0^t \theta_s^n\, (Q_s^n - K_s^n)^+\, ds, \tag{2.4b}$$

$$\langle M^{n,B}\rangle_t = \int_0^t \mu_s^n\, (Q_s^n \wedge K_s^n)\, ds. \tag{2.4c}$$

In addition, these locally square integrable martingales are pairwise orthogonal, i.e., their mutual predictable characteristics are equal to zero:

$$\langle M^{n,A}, M^{n,R}\rangle_t = \langle M^{n,A}, M^{n,B}\rangle_t = \langle M^{n,R}, M^{n,B}\rangle_t = 0\,. \tag{2.5}$$

Proof According to the definition, $M^{n,A}$ is an $\mathbf{F}^n$-martingale. Since $\mathbf{E}(M_t^{n,A})^2 = \int_0^t \lambda_s^n\, ds < \infty$, it is a locally square integrable martingale. One easily checks that $\big((M_t^{n,A})^2 - \int_0^t \lambda_s^n\, ds,\, t \in \mathbb{R}_+\big)$ is a martingale. Similarly, the processes $(H_t^{n,R,i},\, t \in \mathbb{R}_+)$ and $(H_t^{n,B,i},\, t \in \mathbb{R}_+)$, where $H_t^{n,R,i} = R_t^{n,i} - \int_0^t \theta_s^n\, ds$ and $H_t^{n,B,i} = B_t^{n,i} - \int_0^t \mu_s^n\, ds$, are pairwise orthogonal $\mathbf{F}^n$-locally square integrable martingales with predictable quadratic variation processes $\langle H^{n,R,i}\rangle_t = \int_0^t \theta_s^n\, ds$ and $\langle H^{n,B,i}\rangle_t = \int_0^t \mu_s^n\, ds$,respectively.

One can write by (2.2b) and (2.2c) that

$$M_t^{n,R} = \sum_{i=1}^{\infty} M_t^{n,R,i}$$

and

$$M_t^{n,B} = \sum_{i=1}^{\infty} M_t^{n,B,i},$$

where

$$M_t^{n,R,i} = \int_0^t \mathbf{1}_{\{Q_{s-}^n \ge K_{s-}^n + i\}}\, (dR_s^{n,i} - \theta_s^n\, ds)$$

and

$$M_t^{n,B,i} = \int_0^t \mathbf{1}_{\{Q_{s-}^n \wedge K_{s-}^n \ge i\}}\, (dB_s^{n,i} - \mu_s^n\, ds)\,.$$

As stochastic integrals with respect to locally square integrable martingales, the processes $M^{n,R,i} = (M_t^{n,R,i},\, t \in \mathbb{R}_+)$ and $M^{n,B,i} = (M_t^{n,B,i},\, t \in \mathbb{R}_+)$ are $\mathbf{F}^n$-locally square integrable martingales. Their mutual predictable characteristics are given by $\langle M^{n,R,i}, M^{n,R,i'}\rangle_t = \int_0^t \mathbf{1}_{\{Q^n_{s-} \ge K^n_{s-} + i \vee i'\}}\, d\langle H^{n,R,i}, H^{n,R,i'}\rangle_s$, $\langle M^{n,B,i}, M^{n,B,i'}\rangle_t = \int_0^t \mathbf{1}_{\{Q^n_{s-} \wedge K^n_{s-} \ge i \vee i'\}}\, d\langle H^{n,B,i}, H^{n,B,i'}\rangle_s$, and $\langle M^{n,R,i}, M^{n,B,i'}\rangle_t = \int_0^t \mathbf{1}_{\{Q^n_{s-} \ge K^n_{s-} + i\}}\, \mathbf{1}_{\{Q^n_{s-} \wedge K^n_{s-} \ge i'\}}\, d\langle H^{n,R,i}, H^{n,B,i'}\rangle_s$. Therefore, the locally square integrable martingales $M^{n,R,i}$ and $M^{n,B,i}$ are pairwise orthogonal with respective predictable quadratic variation processes $(\int_0^t \mathbf{1}_{\{Q^n_s \ge K^n_s + i\}}\, \theta^n_s\, ds,\, t \in \mathbb{R}_+)$ and $(\int_0^t \mathbf{1}_{\{Q^n_s \wedge K^n_s \ge i\}}\, \mu^n_s\, ds,\, t \in \mathbb{R}_+)$. The stopping times $\tau^n_k = \inf\{t \in \mathbb{R}_+ : Q^n_t \ge k\}$, where $k \in \mathbb{N}$, are common localising times for these locally square integrable martingales and $\sum_{i=1}^\infty \mathbf{E}(M^{n,R,i}_{t \wedge \tau^n_k})^2 < \infty$ and $\sum_{i=1}^\infty \mathbf{E}(M^{n,B,i}_{t \wedge \tau^n_k})^2 < \infty$. It follows that, when stopped at τ^n_k, the processes $M^{n,R}$ and $M^{n,B}$ are square integrable martingales, so they are locally square integrable martingales with respective predictable quadratic variation processes $(\sum_{i=1}^\infty \int_0^t \mathbf{1}_{\{Q^n_s \ge K^n_s + i\}}\, \theta^n_s\, ds,\, t \in \mathbb{R}_+)$ and $(\sum_{i=1}^\infty \int_0^t \mathbf{1}_{\{Q^n_s \wedge K^n_s \ge i\}}\, \mu^n_s\, ds,\, t \in \mathbb{R}_+)$. The fact that the locally square integrable martingales $M^{n,A}$, $M^{n,R}$, and $M^{n,B}$ are pairwise orthogonal follows since those processes have $\mathbf{P}$-a.s. pairwise disjoint jumps.

The next theorem establishes a fluid-scale limit. In the rest of the paper, I will assume as fixed $\mathbb{R}_+$-valued locally integrable functions $(\lambda_t,\, t \in \mathbb{R}_+)$, $(\mu_t,\, t \in \mathbb{R}_+)$, and $(\theta_t,\, t \in \mathbb{R}_+)$, and an $\mathbb{R}_+$-valued Lebesgue measurable function $(\kappa_t,\, t \in \mathbb{R}_+)$. Given an $\mathbb{R}_+$-valued random variable q_0, let q_t be defined by the equation

$$q_t = q_0 + \int_0^t \lambda_s\, ds - \int_0^t \theta_s (q_s - \kappa_s)^+\, ds - \int_0^t \mu_s\, (q_s \wedge \kappa_s)\, ds. \tag{2.7}$$

The Lipshitz continuity of $(x - \kappa_s)^+$ and of $x \wedge \kappa_s$ in x ensures that the equation has a unique solution.

Theorem 1 *Suppose that, as $n \to \infty$, $\int_0^t \lambda^n_s/n\, ds \to \int_0^t \lambda_s\, ds$ for all t, that $\mu^n_t \to \mu_t$ uniformly on bounded intervals, that $\theta^n_t \to \theta_t$ uniformly on bounded intervals, and that $K^n_t/n \to \kappa_t$ for all t. If the random variables Q^n_0/n converge in distribution to a random variable q_0 as $n \to \infty$, then the processes $(Q^n_t/n,\, t \in \mathbb{R}_+)$ converge in distribution in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ to the process $(q_t,\, t \in \mathbb{R}_+)$. In particular, if q_0 is deterministic, then for all $L > 0$ and $\epsilon > 0$,*

$$\lim_{n \to \infty} \mathbf{P}\big(\sup_{t \in [0,L]} \big|\frac{Q^n_t}{n} - q_t\big| > \epsilon\big) = 0\,.$$

Proof Let me first assume that q_0 is deterministic so that the Q^n_0/n converge to q_0 in probability. I prove that

$$\lim_{n \to \infty} \mathbf{P}\big(\frac{1}{n} \sup_{t \in [0,L]} |M^{n,i}_t| > \epsilon\big) = 0 \tag{2.8}$$

for $i = A, R, B$, where $L > 0$ and $\epsilon > 0$ are otherwise arbitrary. The Lénglart-Rebolledo inequality, see, e.g., Liptser and Shiryayev [8, Theorem 1.9.3], implies that it suffices to prove that, for $t \in \mathbb{R}_+$,

$$\lim_{n \to \infty} \mathbf{P}\big(\frac{1}{n^2}\, \langle M^{n,i}\rangle_t > \epsilon\big) = 0\,. \tag{2.9}$$

The validity of (2.9) for $i = A$ follows by (2.4a) and the hypothesis that $\int_0^t \big(\lambda^n_s/n - \lambda_s\big)\, ds \to 0$. As a consequence of (2.8) for $i = A$, I have that the A^n_t/n converge in probability to $\int_0^t \lambda_s\, ds$ as $n \to \infty$ uniformly on bounded intervals. In order to establish (2.9) for $i = R$, I note that by (2.1) $Q^n_t/n \le Q^n_0/n + A^n_t/n$. Since the latter quantities converge in probability uniformly on bounded intervals to $q_0 + \int_0^t \lambda_s\, ds$ as $n \to \infty$, it follows that $\limsup_{n \to \infty} \mathbf{P}(\sup_{s \in [0,t]} Q^n_s/n > q_0 + \int_0^t \lambda_s\, ds + 1) \le \limsup_{n \to \infty} \mathbf{P}(Q^n_0/n + A^n_t/n > q_0 + \int_0^t \lambda_s\, ds + 1) = 0$. If n is such that $\sup_{s \in [0,t]} |\theta^n_s - \theta_s| \le 1$, then by (2.4b), $\mathbf{P}(\langle M^{n,R}\rangle_t/n^2 > \epsilon) \le \mathbf{P}(\sup_{s \in [0,t]}(Q^n_s/n) \int_0^t (\theta_s + 1)\, ds > n\epsilon)$, which implies (2.9) for $i = R$. The case $i = B$ is treated similarly. The limits in (2.8) have been proved.

By (2.3) and (2.7),

$$
\begin{aligned}
|\frac{Q^n_t}{n}-q_t| \le |\frac{Q^n_0}{n}-q_0| &+ |\int\limits_0^t \frac{\lambda^n_s}{n}\,ds-\int\limits_0^t\lambda_s\,ds| + \int\limits_0^t(\theta^n_s+\mu^n_s)|\frac{Q^n_s}{n}-q_s|\,ds\\
&+|\int\limits_0^t \mu^n_s\,\bigl(q_s\wedge\frac{K^n_s}{n}\bigr)\,ds-\int\limits_0^t\mu_s\,(q_s\wedge\kappa_s)\,ds|+|\int\limits_0^t\theta^n_s(q_s-\frac{K^n_s}{n})^+\,ds-\int\limits_0^t\theta_s(q_s-\kappa_s)^+\,ds|\\
&+\frac{1}{n}\,|M^{n,A}_t|+\frac{1}{n}\,|M^{n,R}_t|+\frac{1}{n}\,|M^{n,B}_t|\,.
\end{aligned}
$$

By Gronwall's inequality, see, e.g., p.498 in Ethier and Kurtz [1], for $L>0$,

$$
\begin{aligned}
\sup_{t\in[0,L]}|\frac{Q^n_t}{n}-q_t| \le \bigl(|\frac{Q^n_0}{n}-q_0| &+ \sup_{t\in[0,L]}|\int\limits_0^t\frac{\lambda^n_s}{n}\,ds-\int\limits_0^t\lambda_s\,ds|\\
&+\sup_{t\in[0,L]}|\int\limits_0^t\mu^n_s\,\bigl(q_s\wedge\frac{K^n_s}{n}\bigr)\,ds-\int\limits_0^t\mu_s\,(q_s\wedge\kappa_s)\,ds|\\
&+\sup_{t\in[0,L]}|\int\limits_0^t\theta^n_s(q_s-\frac{K^n_s}{n})^+\,ds-\int\limits_0^t\theta_s(q_s-\kappa_s)^+\,ds|+\frac{1}{n}\sup_{t\in[0,L]}|M^{n,A}_t|\\
&+\frac{1}{n}\sup_{t\in[0,L]}|M^{n,R}_t|+\frac{1}{n}\sup_{t\in[0,L]}|M^{n,B}_t|\bigr)e^{\int_0^L(\theta^n_t+\mu^n_t)\,dt}\,.
\end{aligned}
$$

By (2.8) and the hypotheses, the righthand side tends in probability to zero as $n\to\infty$.

Suppose now that q_0 is random. Let $\mathbf{\Theta}^n_x$ denote the distribution on $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ of $(Q^n_t/n,\,t\in\mathbb{R}_+)$ provided that $Q^n_0/n=x\in\varSigma^n$, where $\varSigma^n=\{0,1/n,2/n,\ldots\}$, let $\mathbf{\Xi}^n$ represent the distribution of Q^n_0/n, and let $\mathbf{\Xi}$ represent the distribution of q_0. By the independence assumptions, it suffices to prove that, for a bounded continuous function f on $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$,

$$
\lim_{n\to\infty}\int\limits_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})\times\varSigma^n} f(z)\,\mathbf{\Theta}^n_x(dz)\,\mathbf{\Xi}^n(dx)=\int\limits_{\mathbb{R}_+}f(q(x))\,\mathbf{\Xi}(dx)\,, \tag{2.10}
$$

where $q(x)=(q_t,\,t\in\mathbb{R}_+)$ is defined by (2.7) with $q_0=x$. By the part just proved, if $x^n\to x$, where $x^n\in\varSigma^n$, and $x\in\mathbb{R}_+$, then the $\mathbf{\Theta}^n_{x^n}$ weakly converge to the Dirac measure at $q(x)$, so

$$
\lim_{n\to\infty}\int\limits_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})} f(z)\,\mathbf{\Theta}^n_{x^n}(dz)=f(q(x))\,. \tag{2.11}
$$

Given $x\in\mathbb{R}_+$, let $g(x)=f(q(x))$ and $g^n(x)=\int_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})}f(z)\,\mathbf{\Theta}^n_{r(x)}(dz)$, where $r(x)$ represents the element of $\varSigma^n$ which is closest to x on the left. By (2.11), if $x^n\to x$, where $x^n\in\mathbb{R}_+$, then $g^n(x^n)\to g(x)$. The weak convergence of the $\mathbf{\Xi}^n$ to $\mathbf{\Xi}$ implies that

$$
\lim_{n\to\infty}\int\limits_{\mathbb{R}_+}g^n(x)\,\mathbf{\Xi}^n(dx)=\int\limits_{\mathbb{R}_+}g(x)\,\mathbf{\Xi}(dx).
$$

Since $\mathbf{\Xi}^n(\varSigma^n)=1$, I have that $\int_{\mathbb{R}_+}g^n(x)\,\mathbf{\Xi}^n(dx)=\int_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})\times\varSigma^n}f(z)\,\mathbf{\Theta}^n_x(dz)\,\mathbf{\Xi}^n(dx)$, so (2.10) follows.

Corollary 1 *Suppose that the hypotheses of Theorem 1 hold where q_0 is deterministic. Then the processes $M^{n,A}/\sqrt{n}=(M^{n,A}_t/\sqrt{n},\,t\in\mathbb{R}_+)$, $M^{n,R}/\sqrt{n}=(M^{n,R}_t/\sqrt{n},\,t\in\mathbb{R}_+)$, and $M^{n,B}/\sqrt{n}=(M^{n,B}_t/\sqrt{n},\,t\in\mathbb{R}_+)$*

jointly converge in distribution in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ *to the respective processes* $M^A = (M^A_t,\, t \in \mathbb{R}_+)$, $M^R = (M^R_t,\, t \in \mathbb{R}_+)$, *and* $M^B = (M^B_t,\, t \in \mathbb{R}_+)$, *defined as follows:*

$$M^A_t = \int_0^t \sqrt{\lambda_s}\, dW^A_s,$$
$$M^R_t = \int_0^t \sqrt{\theta_s (q_s - \kappa_s)^+}\, dW^R_s,$$
$$M^B_t = \int_0^t \sqrt{\mu_s\,(q_s \wedge \kappa_s)}\, dW^B_s,$$

where $W^A = (W^A_t,\, t \in \mathbb{R}_+)$, $W^R = (W^R_t,\, t \in \mathbb{R}_+)$, *and* $W^B = (W^B_t,\, t \in \mathbb{R}_+)$ *are independent standard Wiener processes.*

Proof The processes $M^{n,A}/\sqrt{n}$, $M^{n,R}/\sqrt{n}$, and $M^{n,B}/\sqrt{n}$ are $\mathbf{F}^n$-locally square integrable martingales. By (2.4a), (2.4b), (2.4c), and (2.5) they are mutually orthogonal and their respective predictable quadratic variation processes are given by

$$\langle \frac{M^{n,A}}{\sqrt{n}} \rangle_t = \int_0^t \frac{\lambda^n_s}{n}\, ds,$$
$$\langle \frac{M^{n,R}}{\sqrt{n}} \rangle_t = \int_0^t \theta^n_s \big(\frac{Q^n_s}{n} - \frac{K^n_s}{n}\big)^+\, ds,$$
$$\langle \frac{M^{n,B}}{\sqrt{n}} \rangle_t = \int_0^t \mu^n_s\, \big(\frac{Q^n_s}{n} \wedge \frac{K^n_s}{n}\big)\, ds.$$

By Theorem 1 and the hypotheses, the random variables on the right converge in probability to the functions $\int_0^t \lambda_s\, ds$, $\int_0^t \theta_s (q_s - \kappa_s)^+\, ds$, and $\int_0^t \mu_s\,(q_s \wedge \kappa_s)\, ds$, respectively. (Actually the first convergence is deterministic.) I also have by (2.2a), (2.2b), and (2.2c) that the jumps of the processes $M^{n,A}/\sqrt{n}$, $M^{n,R}/\sqrt{n}$, and $M^{n,B}/\sqrt{n}$ are not greater than $1/\sqrt{n}$. The proof is finished by an application of Theorem 7.1.4 in Liptser and Shiryayev [8].

Let me introduce

$$\alpha^n_t = \sqrt{n}\big(\frac{\lambda^n_t}{n} - \lambda_t\big), \tag{2.12a}$$
$$\beta^n_t = \sqrt{n}\big(\mu^n_t - \mu_t\big), \tag{2.12b}$$
$$\gamma^n_t = \sqrt{n}\big(\theta^n_t - \theta_t\big), \tag{2.12c}$$

and

$$\delta^n_t = \sqrt{n}\big(\frac{K^n_t}{n} - \kappa_t\big)\,. \tag{2.12d}$$

Let processes $X^n = (X^n_t,\, t \in \mathbb{R}_+)$ be defined by

$$X^n_t = \sqrt{n}\big(\frac{Q^n_t}{n} - q_t\big)\,. \tag{2.13}$$

By (2.3) and (2.7), I can write

$$X_t^n = X_0^n + \int_0^t \alpha_s^n\,ds - \int_0^t \theta_s^n\Big(\big(X_s^n + \sqrt{n}q_s - (\delta_s^n + \sqrt{n}\kappa_s)\big)^+ - \sqrt{n}(q_s-\kappa_s)^+\Big)\,ds$$
$$- \int_0^t \mu_s^n\Big(\big(X_s^n + \sqrt{n}q_s\big)\wedge\big(\delta_s^n + \sqrt{n}\kappa_s\big) - \sqrt{n}\,(q_s\wedge\kappa_s)\Big)\,ds - \int_0^t \gamma_s^n(q_s-\kappa_s)^+\,ds$$
$$- \int_0^t \beta_s^n\,(q_s\wedge\kappa_s)\,ds + \frac{1}{\sqrt{n}}\,M_t^{n,A} - \frac{1}{\sqrt{n}}\,M_t^{n,R} - \frac{1}{\sqrt{n}}\,M_t^{n,B}\,. \quad (2.14)$$

In the rest of the paper, $(\alpha_t,\,t\in\mathbb{R}_+)$, $(\beta_t,\,t\in\mathbb{R}_+)$, and $(\gamma_t,\,t\in\mathbb{R}_+)$ represent locally integrable functions and $(\delta_t,\,t\in\mathbb{R}_+)$ represents a locally bounded Lebesgue measurable function.

The following theorem yields a diffusion-scale limit.

Theorem 2 *Let the hypotheses of Theorem 1 hold where $q_0\in\mathbb{R}_+$ is deterministic. Suppose that $\int_0^t \alpha_s^n\,ds \to \int_0^t \alpha_s\,ds$, $\beta_t^n\to\beta_t$, $\gamma_t^n\to\gamma_t$, and $\delta_t^n\to\delta_t$ uniformly on bounded intervals as $n\to\infty$, and that the random variables X_0^n converge in distribution to a random variable X_0 as $n\to\infty$. Then the processes X^n converge in distribution in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ to the process $X=(X_t\,,t\in\mathbb{R}_+)$ that is the solution of the equation*

$$X_t = X_0 + \int_0^t \big(\alpha_s - \gamma_s(q_s-\kappa_s)^+ - \beta_s(q_s\wedge\kappa_s)\big)\,ds - \int_0^t \theta_s\big(\mathbf{1}_{\{q_s>\kappa_s\}}\,(X_s-\delta_s) + \mathbf{1}_{\{q_s=\kappa_s\}}\,(X_s-\delta_s)^+\big)\,ds$$
$$- \int_0^t \mu_s\big(\mathbf{1}_{\{q_s<\kappa_s\}}\,X_s + \mathbf{1}_{\{q_s=\kappa_s\}}\,(X_s\wedge\delta_s) + \mathbf{1}_{\{q_s>\kappa_s\}}\,\delta_s\big)\,ds + \int_0^t \sqrt{\lambda_s + \theta_s(q_s-\kappa_s)^+ + \mu_s\,(q_s\wedge\kappa_s)}\,dW_s,$$

where $W=(W_t,\,t\in\mathbb{R}_+)$ is a standard Wiener process and W and X_0 are independent.

Proof The equation for X has a unique strong solution by the fact that the infinitesimal drift coefficients are Lipshitz continuous, the functions $(\lambda_s,\,s\in\mathbb{R}_+)$, $(\mu_s,\,s\in\mathbb{R}_+)$, $(\theta_s,\,s\in\mathbb{R}_+)$, $(\alpha_s,\,s\in\mathbb{R}_+)$, $(\beta_s,\,s\in\mathbb{R}_+)$, and $(\gamma_s,\,s\in\mathbb{R}_+)$ are locally integrable, and the function $(\delta_s,\,s\in\mathbb{R}_+)$ is locally bounded, see, e.g., Ikeda and Watanabe [6].

Let me first consider the case of deterministic X_0^n, so $X_0^n = x^n\in S^n$, where S^n represents the set of numbers of the form $\sqrt{n}(m/n-q_0)$ for $m\in\mathbb{Z}_+$, and $x^n\to x\in\mathbb{R}$ as $n\to\infty$. By (2.14),

$$|X_t^n| \le |x^n| + |\int_0^t \alpha_s^n\,ds| + \int_0^t (\theta_s^n+\mu_s^n)|X_s^n|\,ds + \int_0^t (\theta_s^n+\mu_s^n)|\delta_s^n|\,ds$$
$$+ \int_0^t |\gamma_s^n|q_s\,ds + \int_0^t |\beta_s^n|\,q_s\,ds + \frac{1}{\sqrt{n}}\,|M_t^{n,A}| + \frac{1}{\sqrt{n}}\,|M_t^{n,R}| + \frac{1}{\sqrt{n}}\,|M_t^{n,B}|\,.$$

Gronwall's inequality, the hypotheses of Theorem 2, and Corollary 1 imply that for $L>0$

$$\lim_{r\to\infty}\limsup_{n\to\infty}\mathbf{P}(\sup_{t\in[0,L]}|X_t^n|>r) = 0\,. \quad (2.15)$$

Also, for $s\le t$,

$$|X_t^n - X_s^n| \le |\int_s^t \alpha_u^n\,du| + \int_s^t (\theta_u^n+\mu_u^n)|X_u^n|\,du + \int_s^t (\theta_u^n+\mu_u^n)|\delta_u^n|\,du$$
$$+ \int_s^t |\gamma_u^n|q_u\,du + \int_s^t |\beta_u^n|\,q_u\,du + \frac{1}{\sqrt{n}}\,|M_t^{n,A}-M_s^{n,A}| + \frac{1}{\sqrt{n}}\,|M_t^{n,R}-M_s^{n,R}| + \frac{1}{\sqrt{n}}\,|M_t^{n,B}-M_s^{n,B}|\,.$$

Given $L>0$, $\eta>0$, and $r>0$,

$$\mathbf{P}\big(\sup_{s,t\in[0,L]:\,|s-t|<\delta}|X_t^n-X_s^n|>\eta\big)\le\mathbf{P}\big(\sup_{t\in[0,L]}|X_t^n|>r\big)\\+\mathbf{P}\Big(\sup_{s,t\in[0,L]:\,|s-t|<\delta}\Big(\Big|\int_s^t\alpha_u^n\,du\Big|+r\int_s^t(\theta_u^n+\mu_u^n)\,du+\int_s^t(\theta_u^n+\mu_u^n)|\delta_u^n|\,du\\+\int_s^t|\gamma_u^n|q_u\,du+\int_s^t|\beta_u^n|\,q_u\,du+\frac{1}{\sqrt{n}}\,|M_t^{n,A}-M_s^{n,A}|\\+\frac{1}{\sqrt{n}}\,|M_t^{n,R}-M_s^{n,R}|+\frac{1}{\sqrt{n}}\,|M_t^{n,B}-M_s^{n,B}|\Big)>\eta\Big)\,.$$

Hence, by Corollary 1 and the hypotheses,

$$\limsup_{n\to\infty}\mathbf{P}\big(\sup_{s,t\in[0,L]:\,|s-t|<\delta}|X_t^n-X_s^n|>\eta\big)\le\limsup_{n\to\infty}\mathbf{P}\big(\sup_{t\in[0,L]}|X_t^n|>r\big)\\+\mathbf{P}\Big(\sup_{s,t\in[0,L]:\,|s-t|<\delta}\Big(\Big|\int_s^t\alpha_u\,du\Big|+r\int_s^t(\theta_u+\mu_u)\,du+\int_s^t(\theta_u+\mu_u)|\delta_u|\,du\\+\int_s^t|\gamma_u|q_u\,du+\int_s^t|\beta_u|\,q_u\,du+|M_t^A-M_s^A|+|M_t^R-M_s^R|+|M_t^B-M_s^B|\Big)>\frac{\eta}{2}\Big)\,.$$

By the continuity of the processes M^A, M^B, and M^R, and absolute continuity of the Lebesgue integral, the limit of the second probability on the right, as $\delta\to0$, equals zero, so

$$\limsup_{\delta\to0}\limsup_{n\to\infty}\mathbf{P}\big(\sup_{s,t\in[0,L]:\,|s-t|<\delta}|X_t^n-X_s^n|>\eta\big)\le\limsup_{n\to\infty}\mathbf{P}\big(\sup_{t\in[0,L]}|X_t^n|>r\big)\,.$$

By (2.15), the righthand side can be made arbitrarily small by choosing r great enough. Therefore,

$$\lim_{\delta\to0}\limsup_{n\to\infty}\mathbf{P}\big(\sup_{s,t\in[0,L]:\,|s-t|<\delta}|X_t^n-X_s^n|>\eta\big)=0\,.$$

It follows that the sequence X^n is $\mathbb{C}$-tight, i.e., it is tight for convergence in distribution in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$, and all limit points are continuous-path processes. Let $\tilde X=(\tilde X_t,\,t\in\mathbb{R}_+)$ represent a subsequential limit of the X^n.

Let me note that if a sequence of functions $(x_t^n,\,t\in\mathbb{R}_+)$ from $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ converges for Skorohod's J_1-topology to a continuous function $(x_t,\,t\in\mathbb{R}_+)$ as $n\to\infty$, then

$$\int_0^t\theta_s^n\big(\big(x_s^n+\sqrt{n}q_s-(\delta_s^n+\sqrt{n}\kappa_s)\big)^+-\sqrt{n}(q_s-\kappa_s)^+\big)\,ds\to\int_0^t\theta_s\big(\mathbf{1}_{\{q_s>\kappa_s\}}\,(x_s-\delta_s)+\mathbf{1}_{\{q_s=\kappa_s\}}\,(x_s-\delta_s)^+\big)\,ds \tag{2.16}$$

and

$$\int_0^t\mu_s^n\big(\big(x_s^n+\sqrt{n}q_s\big)\wedge\big(\delta_s^n+\sqrt{n}\kappa_s\big)-\sqrt{n}\,(q_s\wedge\kappa_s)\big)\,ds\to\int_0^t\mu_s\big(\mathbf{1}_{\{q_s<\kappa_s\}}\,x_s+\mathbf{1}_{\{q_s=\kappa_s\}}\,(x_s\wedge\delta_s)+\mathbf{1}_{\{q_s>\kappa_s\}}\,\delta_s\big)\,ds\,. \tag{2.17}$$

To see (2.16), one could first note that $\theta_s^n\big(\big(x_s^n+\sqrt{n}q_s-(\delta_s^n+\sqrt{n}\kappa_s)\big)^+-\sqrt{n}(q_s-\kappa_s)^+\big)\to\theta_s\big(\mathbf{1}_{\{q_s>\kappa_s\}}\,(x_s-\delta_s)+\mathbf{1}_{\{q_s=\kappa_s\}}\,(x_s-\delta_s)^+\big)$ for each s, for if $q_s>\kappa_s$, then $\big(x_s^n+\sqrt{n}q_s-(\delta_s^n+\sqrt{n}\kappa_s)\big)^+-\sqrt{n}(q_s-\kappa_s)^+\big)=\big(x_s^n+\sqrt{n}q_s-(\delta_s^n+\sqrt{n}\kappa_s)\big)-\sqrt{n}(q_s-\kappa_s)\big)=x_s^n-\delta_s^n$ for all n great enough, if $q_s=\kappa_s$, then $\big(x_s^n+\sqrt{n}q_s-(\delta_s^n+\sqrt{n}\kappa_s)\big)^+-\sqrt{n}(q_s-\kappa_s)^+\big)=(x_s^n-\delta_s^n)^+$, and if $q_s<\kappa_s$, then $\big(x_s^n+\sqrt{n}q_s-(\delta_s^n+\sqrt{n}\kappa_s)\big)^+-\sqrt{n}(q_s-\kappa_s)^+\big)=0$

for all n great enough. Since $|\theta^n_s\big(\big(x^n_s+\sqrt{n}q_s-(\delta^n_s+\sqrt{n}\kappa_s)\big)^+-\sqrt{n}(q_s-\kappa_s)^+\big)|\le\theta^n_s|x^n_s-\delta^n_s|$, the convergence in (2.16) follows by Lebesgue's dominated convergence theorem. The argument for (2.17) is similar.

On recalling Corollary 1, I conclude from (2.14) and the continuous mapping principle that $\tilde{X}$ must satisfy the equation

$$\tilde{X}_t = x+\int\limits_0^t \alpha_s\,ds-\int\limits_0^t\theta_s\big(\mathbf{1}_{\{q_s>\kappa_s\}}(\tilde{X}_s-\delta_s)+\mathbf{1}_{\{q_s=\kappa_s\}}(\tilde{X}_s-\delta_s)^+\big)\,ds-\int\limits_0^t\gamma_s(q_s-\kappa_s)^+\,ds$$
$$-\int\limits_0^t\mu_s\big(\mathbf{1}_{\{q_s<\kappa_s\}}\tilde{X}_s+\mathbf{1}_{\{q_s=\kappa_s\}}(\tilde{X}_s\wedge\delta_s)+\mathbf{1}_{\{q_s>\kappa_s\}}\delta_s\big)\,ds-\int\limits_0^t\beta_s(q_s\wedge\kappa_s)\,ds$$
$$+\int\limits_0^t\sqrt{\lambda_s+\theta_s(q_s-\kappa_s)^++\mu_s\,(q_s\wedge\kappa_s)}\,d\tilde{W}_s\,,$$

where $(\tilde{W}_t,\,t\in\mathbb{R}_+)$ is a standard Wiener process. Since the latter equation has a unique solution, $\tilde{X}$ coincides in law with X, so the X^n converge in distribution to X.

I will now consider the case of general X^n_0. The argument is similar to the one used in the proof of Theorem 1. Let $\mathbf{\Phi}^n_x$ denote the distribution on $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ of X^n provided that $X^n_0=x\in S^n$, let $\mathbf{\Phi}_x$ denote the distribution on $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ of X provided that $X_0=x\in\mathbb{R}$, let $\mathbf{\Psi}^n$ denote the distribution of X^n_0, and let $\mathbf{\Psi}$ denote the distribution of X_0. By the independence assumptions, it suffices to prove that, for a bounded continuous function f on $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$,

$$\lim_{n\to\infty}\int\limits_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})\times S^n} f(z)\,\mathbf{\Phi}^n_x(dz)\,\mathbf{\Psi}^n(dx)=\int\limits_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})\times\mathbb{R}} f(z)\,\mathbf{\Phi}_x(dz)\,\mathbf{\Psi}(dx)\,. \tag{2.18}$$

By the part just proved, if $x^n\to x$, where $x^n\in S^n$ and $x\in\mathbb{R}$, then the $\mathbf{\Phi}^n_{x^n}$ weakly converge to $\mathbf{\Phi}_x$, so

$$\lim_{n\to\infty}\int\limits_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})} f(z)\,\mathbf{\Phi}^n_{x^n}(dz)=\int\limits_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})} f(z)\,\mathbf{\Phi}_x(dz)\,. \tag{2.19}$$

Given $x\in\mathbb{R}$, let $g(x)=\int_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})}f(z)\,\mathbf{\Phi}_x(dz)$ and $g^n(x)=\int_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})}f(z)\,\mathbf{\Phi}^n_{r(x)}(dz)$, where $r(x)$ represents the element of S^n which is closest to x on the left. By (2.19), if $x^n\to x$, where $x^n\in\mathbb{R}$ and $x\in\mathbb{R}$, then $g^n(x^n)\to g(x)$. The weak convergence of the $\mathbf{\Psi}^n$ to $\mathbf{\Psi}$ implies that

$$\lim_{n\to\infty}\int\limits_{\mathbb{R}} g^n(x)\,\mathbf{\Psi}^n(dx)=\int\limits_{\mathbb{R}} g(x)\,\mathbf{\Psi}(dx).$$

Since $\mathbf{\Psi}^n(S^n)=1$, I have that $\int_{\mathbb{R}}g^n(x)\,\mathbf{\Psi}^n(dx)=\int_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})\times S^n}f(z)\,\mathbf{\Phi}^n_x(dz)\,\mathbf{\Psi}^n(dx)$, so (2.18) follows.

Remark 1 The assertions of Theorem 1 and Theorem 2 are contained in Theorem 2.2 and Theorem 2.3, respectively, in Mandelbaum, Massey, and Reiman [10], albeit under slightly stronger hypotheses. However, the proof of Lemma 9.3 there depends on the erroneous claim that if a sequence of nonnegative random variables defined on the same probability space is tight, then it has a finite limit superior a.s. There are also problems with establishing the martingale property in the proof of Lemma 9.1.

3 Convergence of the periodic queue lengths

In this section, I will assume that the functions λ^n_t, μ^n_t, θ^n_t, K^n_t, λ_t, μ_t, θ_t, and κ_t are T-periodic, where $T>0$. I will also assume that

$$\int\limits_0^T\lambda_s\,ds>0 \tag{3.1}$$

and

$$\int_0^T (\mu_s \wedge \theta_s)\, ds > 0\,. \tag{3.2}$$

In the long term, one expects a periodic pattern to emerge for the number of customers. The next lemma confirms that to be the case. Let $Q^{n,\ell} = (Q^n_{\ell T+t},\, t \in \mathbb{R}_+)$, where $\ell \in \mathbb{Z}_+$. The sequence $\{Q^{n,\ell},\, \ell \in \mathbb{Z}_+\}$ is a discrete-time homogeneous Markov process with values in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$.

Lemma 2 *Suppose that $\int_0^T (\mu^n_s \wedge \theta^n_s)\, ds > 0$. As $\ell \to \infty$, given an arbitrarily distributed Q^n_0, the sequence of the distributions of the processes $Q^{n,\ell}$ converges in the distance of total variation in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ to the distribution of a process $\breve{Q}^n = (\breve{Q}^n_t,\, t \in \mathbb{R}_+)$, which is a unique T-periodic Markov process with the same transition probability function as Q^n . The distribution of $(\breve{Q}^n_t,\, t \in \mathbb{R}_+)$ is a stationary initial distribution for $\{Q^{n,\ell},\, \ell \in \mathbb{Z}_+\}$.*

Remark 2 For the definition of the distance of total variation, see, e.g., p.274 in Jacod and Shiryaev [7].

Proof If $\int_0^T \lambda^n_s\, ds = 0$, then $A^n_t = 0$ for all $t \in \mathbb{R}_+$, so 0 is an absorbing state for Q^n and $Q^n_t \to 0$ in the distance of total variation as $t \to \infty$, so $\breve{Q}^n_t = 0$.

Suppose that $\int_0^T \lambda^n_s\, ds > 0$. Then the sequence $\{Q^n_{\ell T},\, \ell \in \mathbb{Z}_+\}$ is a time-homogeneous, irreducible and aperiodic discrete-time Markov chain. One can show as follows that it converges in the distance of total variation to a unique stationary distribution as $\ell \to \infty$. It suffices to prove that the chain is positive recurrent, which, by Foster's criterion, will follow if, for some $N \in \mathbb{Z}_+$,

$$\mathbf{E}_x Q^n_T \le x - 1 \text{ for all } x \in \{N+1, N+2, \ldots\}\,, \tag{3.3}$$

where $\mathbf{E}_x$ denotes expectation with respect to the probability measure $\mathbf{P}_x$ such that $\mathbf{P}_x(Q^n_0 = x) = 1$, see, e.g., Theorem 11.3.4 on p.265 and Proposition 13.2.4 on p.319 in Meyn and Tweedie [11], or Theorem 2.2.3 on p.29 in Fayolle, Malyshev, and Men'shikov [2]. Since $Q^n_t \le Q^n_0 + A^n_t$ by (2.1), $\mathbf{E}_x Q^n_t \le x + \int_0^t \lambda^n_s\, ds$. By Lemma 1, the processes $M^{n,A}$, $M^{n,R}$, and $M^{n,B}$ are $\mathbf{F}^n$-locally square integrable martingales under $\mathbf{P}_x$ with respective predictable quadratic variation processes $(\int_0^t \lambda^n_s\, ds\,, t \in \mathbb{R}_+)$, $(\int_0^t \theta^n_s\, (Q^n_s - K^n_s)^+\, ds,\, t \in \mathbb{R}_+)$, and $(\int_0^t \mu^n_s\, (Q^n_s \wedge K^n_s)\, ds,\, t \in \mathbb{R}_+)$. Since the latter processes are of finite expectation, $\mathbf{E}(M^{n,A}_t)^2 < \infty$, $\mathbf{E}(M^{n,R}_t)^2 < \infty$, and $\mathbf{E}(M^{n,B}_t)^2 < \infty$. In particular, the processes $M^{n,A}$, $M^{n,R}$, and $M^{n,B}$ are $\mathbf{F}^n$-martingales, so by (2.3),

$$-\mathbf{E}_x Q^n_t \prec -x + \int_0^t (\theta^n_s \vee \mu^n_s) \mathbf{E}_x Q^n_s\, ds\,,$$

which implies by Lemma 5 that

$$\inf_{t \le T} \mathbf{E}_x Q^n_t \ge x e^{-\int_0^T (\mu^n_s \vee \theta^n_s)\, ds}\,.$$

By (2.3),

$$\mathbf{E}_x Q^n_T \le x + \int_0^T \lambda^n_s\, ds - \int_0^T (\mu^n_s \wedge \theta^n_s)\, \mathbf{E}_x Q^n_s\, ds \le x + \int_0^T \lambda^n_s\, ds - x e^{-\int_0^T (\mu^n_s \vee \theta^n_s)\, ds} \int_0^T \mu^n_s \wedge \theta^n_s\, ds\,.$$

Therefore, (3.3) holds if

$$N \ge e^{\int_0^T (\mu^n_s \vee \theta^n_s)\, ds} \frac{1 + \int_0^T \lambda^n_s\, ds}{\int_0^T \mu^n_s \wedge \theta^n_s\, ds}\,.$$

Thus, the distributions of $Q^n_{\ell T}$ converge in the distance of total variation to a limit distribution as $\ell \to \infty$. Since the transition probability function of $Q^{n,\ell}$ is periodic, the finite-dimensional distributions of $Q^{n,\ell}$ converge in the distance of total variation to limit distributions as $\ell \to \infty$. Since the Borel and cylindrical σ-algebras on $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ coincide, it follows that the distributions of the Markov processes $Q^{n,\ell}$ converge in the distance of total variation to the distribution of a process $\breve{Q}^n = (\breve{Q}^n_t,\, t \in \mathbb{R}_+)$, which is a T-periodic Markov process with the same transition probability function as Q^n. Since the limiting distribution of the $Q^n_{\ell T}$ is specified uniquely, the distribution of $\breve{Q}^n$ is specified uniquely. Since the sequence $\{\breve{Q}^n_{\ell T}\,, \ell \in \mathbb{Z}_+\}$ is stationary and the processes $(\breve{Q}^n_{\ell T+t}\,, t \in \mathbb{R}_+)$ are Markov processes with the same transition probability function, it follows that the sequence $\{(\breve{Q}^n_{\ell T+t}\,, t \in \mathbb{R}_+),\, \ell \in \mathbb{Z}_+\}$ is stationary.

Remark 3 Note that the sequence $\{Q^{n,\ell},\, \ell \in \mathbb{Z}_+\}$ is deterministic once the initial condition $(Q^n_t\,, t \in \mathbb{R}_+)$ has been chosen.

My next step is to consider periodic regimes for the deterministic approximation.

Lemma 3 *1. There exists a unique $q_0 \in \mathbb{R}_+$ such that the function $(q_t,\, t \in \mathbb{R}_+)$ defined by equation (2.7) is T-periodic. An arbitrary solution converges to this periodic solution as $t \to \infty$.*

2. If q_0 is a random variable such that the process $(q_t,\, t \in \mathbb{R}_+)$ is T-periodic, then q_0 is deterministic and has the value specified in part 1.

Proof By uniqueness, no two solutions have a point in common. In particular, if $q'_0 > q_0$, then for the corresponding solutions, $q'_t > q_t$ for all $t \in \mathbb{R}_+$. Given q_0, there are three possibilities: either $q_T = q_0$, or $q_T > q_0$, or $q_T < q_0$. If $q_T = q_0$, then the solution starting at q_0 is a periodic solution. Suppose that $q_T > q_0$. Then on taking q_T as a new initial condition, by periodicity, $q_{t+T} > q_t$ for all $t \in \mathbb{R}_+$, so $q_{2T} > q_T$. Continuing on, I obtain an increasing sequence of solutions $(q_{\ell T+t},\, t \in \mathbb{R}_+)$, where $\ell = 0, 1, 2, \ldots$. By part 1(a) of Lemma 6 found in the appendix, $\sup_{t\in\mathbb{R}_+} q_t < \infty$, so there exists a limit of $q_{\ell T+t}$ as $\ell \to \infty$. I denote this limit by $\breve{q}_t$. Since $q_{\ell T} \to \breve{q}_0$ and $q_{\ell T+T} \to \breve{q}_T$, $(\breve{q}_t,\, t \in \mathbb{R}_+)$ is a T-periodic function. By bounded convergence, it is also a solution. If $q_T < q_0$, then $(q_{\ell T+t},\, t \in \mathbb{R}_+)$ is a monotonically decreasing sequence of functions converging to a T-periodic solution.

To show the uniqueness of a T-periodic solution, note that if $(\breve{q}_t,\, t \in \mathbb{R}_+)$ is a T-periodic solution, then $\breve{q}_0 = \breve{q}_T$, so $\int_0^T \lambda_s\, ds = \int_0^T \theta_s(\breve{q}_s - \kappa_s)^+\, ds + \int_0^T \mu_s\, (\breve{q}_s \wedge \kappa_s)\, ds$. Now, if $(q'_t,\, t \in \mathbb{R}_+)$ is a solution with $q'_0 > \breve{q}_0$, then $q'_t > \breve{q}_t$ for all t, so on recalling (3.2),

$$\int_0^T \theta_s(q'_s - \kappa_s)^+\, ds + \int_0^T \mu_s\, (q'_s \wedge \kappa_s)\, ds > \int_0^T \theta_s(\breve{q}_s - \kappa_s)^+\, ds + \int_0^T \mu_s\, (\breve{q}_s \wedge \kappa_s)\, ds = \int_0^T \lambda_s\, ds,$$

which implies that $q'_T < q'_0$. Similarly, if $q'_0 < \breve{q}_0$, then $q'_T > q'_0$. Thus, $(q'_t,\, t \in \mathbb{R}_+)$ is not T-periodic. Part 1 is proved.

Let $(q_t,\, t \in \mathbb{R}_+)$ represent a T-periodic process. The reasoning used to show the uniqueness of a T-periodic solution shows that $|q_T - \breve{q}_0| < |q_0 - \breve{q}_0|$ when $q_0 \neq \breve{q}_0$. Since the distributions of $|q_T - \breve{q}_0|$ and $|q_0 - \breve{q}_0|$ are the same, $q_0 = \breve{q}_0$ a.s.

In what follows, $(\breve{q}_t,\, t \in \mathbb{R}_+)$ represents the T-periodic solution of Lemma 3.

Theorem 3 *Suppose that, as $n \to \infty$, $\int_0^t \lambda^n_s/n\, ds \to \int_0^t \lambda_s\, ds$ for all t, that $\mu^n_t \to \mu_t$ uniformly on bounded intervals, that $\theta^n_t \to \theta_t$ uniformly on bounded intervals, and that $K^n_t/n \to \kappa_t$ for all t. Then, for all $\epsilon > 0$ and $L > 0$,*

$$\lim_{n\to\infty} \mathbf{P}\big(\sup_{t\in[0,L]} \big|\frac{\breve{Q}^n_t}{n} - \breve{q}_t\big| > \epsilon\big) = 0\,.$$

Proof Since $\breve{Q}^n_0$ is a limit in distribution of the Q^n_t as $t \to \infty$, by part 1(b) of Lemma 6 (with $Q^n_0 = 0$), the sequence $\{\breve{Q}^n_0/n,\, n \in \mathbb{N}\}$ is tight. (Note that by (3.2), $\liminf_{n\to\infty} \int_0^T (\mu^n_s \wedge \theta^n_s)\, ds > 0$.) By Theorem 1 and Prohorov's theorem, the sequence of processes $\{(\breve{Q}^n_t/n,\, t \in \mathbb{R}_+),\, n \in \mathbb{N}\}$ is tight and any limit point $(q_t,\, t \in \mathbb{R}_+)$ is the solution of (2.7) for a suitable q_0. Since the processes $(\breve{Q}^n_t/n,\, t \in \mathbb{R}_+)$ are T-periodic, so is the process $(q_t,\, t \in \mathbb{R}_+)$. By Lemma 3, $q_t = \breve{q}_t$ a.s., which concludes the proof.

Let

$$\breve{X}_t^n = \sqrt{n}\bigl(\frac{\breve{Q}_t^n}{n} - \breve{q}_t\bigr)\,. \tag{3.4}$$

The process $(\breve{X}_t^n,\, t \in \mathbb{R}_+)$ is a T-periodic Markov process.

Theorem 4 *Suppose that $\int_0^t \alpha_s^n\, ds \to \int_0^t \alpha_s\, ds$, $\beta_t^n \to \beta_t$, $\gamma_t^n \to \gamma_t$, and $\delta_t^n \to \delta_t$ uniformly on bounded intervals as $n \to \infty$. Then the processes $(\breve{X}_t^n,\, t \in \mathbb{R}_+)$ converge in distribution as $n \to \infty$ to process $(\breve{X}_t,\, t \in \mathbb{R}_+)$, which is a unique T-periodic Markov process satisfying the equation*

$$\breve{X}_t = \breve{X}_0 + \int_0^t \bigl(\alpha_s - \gamma_s(\breve{q}_s - \kappa_s)^+ - \beta_s(\breve{q}_s \wedge \kappa_s)\bigr)\, ds - \int_0^t \theta_s\bigl(\mathbf{1}_{\{\breve{q}_s > \kappa_s\}}\,(\breve{X}_s - \delta_s) + \mathbf{1}_{\{\breve{q}_s = \kappa_s\}}\,(\breve{X}_s - \delta_s)^+\bigr)\, ds$$
$$- \int_0^t \mu_s\bigl(\mathbf{1}_{\{\breve{q}_s < \kappa_s\}}\,\breve{X}_s + \mathbf{1}_{\{\breve{q}_s = \kappa_s\}}\,(\breve{X}_s \wedge \delta_s) + \mathbf{1}_{\{\breve{q}_s > \kappa_s\}}\,\delta_s\bigr)\, ds + \int_0^t \sqrt{\lambda_s + \theta_s(\breve{q}_s - \kappa_s)^+ + \mu_s\,(\breve{q}_s \wedge \kappa_s)}\, d\breve{W}_s,$$

where $(\breve{W}_t,\, t \in \mathbb{R}_+)$ is a standard Wiener process and $\breve{X}_0$ and $(\breve{W}_t,\, t \in \mathbb{R}_+)$ are independent.

Proof By Lemma 2 and Lemma 3, the processes $(X_{\ell T+t}^n,\, t \in \mathbb{R}_+)$, where $Q_0^n = q_0 = 0$, converge in distribution in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ to $(\breve{X}_t^n,\, t \in \mathbb{R}_+)$ as $\ell \to \infty$ and the sequence $\{(\breve{X}_{\ell T+t}^n,\, t \in \mathbb{R}_+),\, \ell \in \mathbb{Z}_+\}$ is stationary. By part 1(c) of Lemma 6, $\lim_{V\to\infty} \limsup_{n\to\infty} \limsup_{t\to\infty} \mathbf{P}(|X_t^n| > V) = 0$. Therefore, the sequence $\{\breve{X}_0^n\,, n \in \mathbb{N}\}$ is tight. By Theorem 2 and Prohorov's theorem, the sequence $\{(\breve{X}_t^n, t \in \mathbb{R}_+)\,, n \in \mathbb{N}\}$ is tight. Let $(\breve{X}_t, t \in \mathbb{R}_+)$ represent a limit point of that sequence for convergence in distribution in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ as $n \to \infty$. As follows by Theorem 2, it satisfies the equation in the statement and is a Markov process. In addition, $\{(\breve{X}_{\ell T+t},\, t \in \mathbb{R}_+),\, \ell \in \mathbb{Z}_+\}$ is a limit point of $\{(\breve{X}_{\ell T+t}^n,\, t \in \mathbb{R}_+),\, \ell \in \mathbb{Z}_+\}$ as $n \to \infty$ for convergence in distribution in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})^{\mathbb{Z}_+}$. Since the sequence $\{(\breve{X}_{\ell T+t}^n,\, t \in \mathbb{R}_+),\, \ell \in \mathbb{Z}_+\}$ is stationary, so is the sequence $\{(\breve{X}_{\ell T+t},\, t \in \mathbb{R}_+),\, \ell \in \mathbb{Z}_+\}$. Hence, $(\breve{X}_t, t \in \mathbb{R}_+)$ is a T-periodic Markov process.

The following coupling argument shows that the distribution of $(\breve{X}_t,\, t \in \mathbb{R}_+)$ is specified uniquely and is the limit of the distributions of processes $(\tilde{X}_{\ell T+t},\, t \in \mathbb{R}_+)$, as $\ell \to \infty$, where the process $\tilde{X} = (\tilde{X}_t,\, t \in \mathbb{R}_+)$ is defined by the equation

$$\tilde{X}_t = x + \int_0^t \bigl(\alpha_s - \gamma_s(\breve{q}_s - \kappa_s)^+ - \beta_s(\breve{q}_s \wedge \kappa_s)\bigr)\, ds - \int_0^t \theta_s\bigl(\mathbf{1}_{\{\breve{q}_s > \kappa_s\}}\,(\tilde{X}_s - \delta_s) + \mathbf{1}_{\{\breve{q}_s = \kappa_s\}}\,(\tilde{X}_s - \delta_s)^+\bigr)\, ds$$
$$- \int_0^t \mu_s\bigl(\mathbf{1}_{\{\breve{q}_s < \kappa_s\}}\,\tilde{X}_s + \mathbf{1}_{\{\breve{q}_s = \kappa_s\}}\,(\tilde{X}_s \wedge \delta_s) + \mathbf{1}_{\{\breve{q}_s > \kappa_s\}}\,\delta_s\bigr)\, ds + \int_0^t \sqrt{\lambda_s + \theta_s(\breve{q}_s - \kappa_s)^+ + \mu_s\,(\breve{q}_s \wedge \kappa_s)}\, d\tilde{W}_s\,,$$

where $x \in \mathbb{R}$ and $(\tilde{W}_t,\, t \in \mathbb{R}_+)$ is a standard Wiener process. Let me consider a process $\tilde{X}' = (\tilde{X}_t',\, t \in \mathbb{R}_+)$ which starts at $y \in \mathbb{R}$ and is driven by the negative of the Wiener process $(\tilde{W}_t,\, t \in \mathbb{R}_+)$ so that

$$\tilde{X}_t' = y + \int_0^t \bigl(\alpha_s - \gamma_s(\breve{q}_s - \kappa_s)^+ - \beta_s(\breve{q}_s \wedge \kappa_s)\bigr)\, ds - \int_0^t \theta_s\bigl(\mathbf{1}_{\{\breve{q}_s > \kappa_s\}}\,(\tilde{X}_s' - \delta_s) + \mathbf{1}_{\{\breve{q}_s = \kappa_s\}}\,(\tilde{X}_s' - \delta_s)^+\bigr)\, ds$$
$$- \int_0^t \mu_s\bigl(\mathbf{1}_{\{\breve{q}_s < \kappa_s\}}\,\tilde{X}_s' + \mathbf{1}_{\{\breve{q}_s = \kappa_s\}}\,(\tilde{X}_s' \wedge \delta_s) + \mathbf{1}_{\{\breve{q}_s > \kappa_s\}}\,\delta_s\bigr)\, ds - \int_0^t \sqrt{\lambda_s + \theta_s(\breve{q}_s - \kappa_s)^+ + \mu_s\,(\breve{q}_s \wedge \kappa_s)}\, d\tilde{W}_s\,.$$

Obviously, the distribution of $\tilde{X}'$ is the same as the distribution of $\tilde{X}$ if the latter were started at y . Assuming that $x > y$, I have that until $\tilde{X}$ and $\tilde{X}'$ meet,

$$\tilde{X}_t - \tilde{X}_t' \leq x - y + 2\int_0^t \sqrt{\lambda_s + \theta_s(\breve{q}_s - \kappa_s)^+ + \mu_s\,(\breve{q}_s \wedge \kappa_s)}\, d\tilde{W}_s\,.$$

For $\tau_{x,y} = \inf\{t : \tilde{X}_t = \tilde{X}'_t\}$,

$$\mathbf{P}(\tau_{x,y} > t) \le \mathbf{P}\big(2 \inf_{u\in[0,t]} \int_0^u \sqrt{\lambda_s + \theta_s(\check{q}_s - \kappa_s)^+ + \mu_s\,(\check{q}_s \wedge \kappa_s)}\, d\tilde{W}_s \ge y - x\big)\,.$$

Let $\Gamma(u) = \inf\{v : \int_0^v \big(\lambda_s + \theta_s(\check{q}_s - \kappa_s)^+ + \mu_s\,(\check{q}_s \wedge \kappa_s)\big)\, ds = u\}$, which is finite by (3.1). Since the processes $(\int_0^u \sqrt{\lambda_s + \theta_s(\check{q}_s - \kappa_s)^+ + \mu_s\,(\check{q}_s \wedge \kappa_s)}\, d\tilde{W}_s,\, u \in \mathbb{R}_+)$ and $(\tilde{W}_{\int_0^u(\lambda_s+\theta_s(\check{q}_s-\kappa_s)^+ + \mu_s\,(\check{q}_s\wedge\kappa_s))\,ds},\, u \in \mathbb{R}_+)$ have the same distribution, $\Gamma^{-1}(u) = \int_0^u(\lambda_s+\theta_s(\check{q}_s-\kappa_s)^+ +\mu_s\,(\check{q}_s\wedge\kappa_s))\,ds$, the random variables $(-\inf_{u\in[0,\Gamma^{-1}(t)]} \tilde{W}_u)$ and $|\tilde{W}_{\Gamma^{-1}(t)}|$ have the same distribution, and the random variables $\tilde{W}_{\Gamma^{-1}(t)}$ and $\sqrt{\Gamma^{-1}(t)}\tilde{W}_1$ have the same distribution, I conclude that

$$\mathbf{P}\big(2 \inf_{u\in[0,t]} \int_0^u \sqrt{\lambda_s + \theta_s(\check{q}_s - \kappa_s)^+ + \mu_s\,(\check{q}_s \wedge \kappa_s)}\, d\tilde{W}_s \ge y - x\big) = \mathbf{P}\big(2 \inf_{u\in[0,\Gamma^{-1}(t)]} \tilde{W}_u \ge y - x\big)$$
$$= \mathbf{P}(2|\tilde{W}_{\Gamma^{-1}(t)}| \le x - y) = \mathbf{P}(2|\tilde{W}_1| \le (x-y)/\sqrt{\Gamma^{-1}(t)}) \to 0 \text{ as } t \to \infty,$$

It follows that $\mathbf{P}(\tau_{x,y} > t) \to 0$ as $t \to \infty$. Furthermore, the latter convergence is uniform over x and y from bounded sets. Therefore, the T-periodic version of $\tilde{X}$ is unique in distribution and one has convergence in the distance of total variation to the distribution of that process from an arbitrary initial distribution. (For a sample argument, let ν denote the distribution of $\check{X}_0$, let $\tilde{\nu}$ denote a probability distribution on $\mathbb{R}$, and let $\nu_{x,\ell}$ denote the distribution of $(\tilde{X}_{\ell T+s},\, s \in \mathbb{R}_+)$ with $\tilde{X}_0 = x$, where $\ell \in \mathbb{Z}_+$. Then, for a bounded measurable function f on $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ and $U \in \mathbb{R}_+$, $|\int_{\mathbb{R}}\int_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})} f(z)\nu_{x,\ell}(dz)\tilde{\nu}(dx) - \int_{\mathbb{R}}\int_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})} f(z)\nu_{x,0}(dz)\nu(dx)| \le 2\sup_{|x|\le U}|\int_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})} f(z)\nu_{x,\ell}(dz) - \int_{\mathbb{D}(\mathbb{R}_+,\mathbb{R})} f(z)\nu_{0,\ell}(dz)| + 2\sup_{x\in\mathbb{R}}|f(x)|(\tilde{\nu}(x : |x| > U) + \nu(x : |x| > U)) \le 2\sup_{x\in\mathbb{R}}|f(x)|\big(\sup_{|x|\le U}\mathbf{P}(\tau_{x,0} > \ell T) + \tilde{\nu}(|x| > U) + \nu(|x| > U)\big)$. The latter expression converges to zero as $\ell \to \infty$ and $U \to \infty$.)

Remark 4 If the functions $(\theta_t,\, t \in \mathbb{R}_+)$ and $(\mu_t,\, t \in \mathbb{R}_+)$ are bounded, then the existence of the periodic version of $\tilde{X}$ can be deduced from Theorem 5.2 on p.90 of Has'minskii [5] (with $V(t,x) = x^2$). I haven't found other results in the literature which directly apply, the sticking point being that the equation coefficients are not differentiable functions of time and space.

4 Convergence of stationary distributions

In this section I will assume constant arrival, service, and abandonment rates, so $\lambda^n_t = \lambda^n \ge 0$, $\theta^n_t = \theta^n > 0$, $\mu^n_t = \mu^n > 0$, $\lambda_t = \lambda \ge 0$, $\theta_t = \theta > 0$, $\mu_t = \mu > 0$, $\alpha_t = \alpha$, $\beta_t = \beta$, and $\gamma_t = \gamma$. The number of servers K^n_t is also assumed to be constant which I will take as the scaling parameter n, so $\kappa_t = 1$. Accordingly, $\delta_t = 0$. The equations for the fluid- and diffusion-scale limits which appear in Theorem 1 and Theorem 2 assume the following form:

$$q_t = q_0 + \lambda t - \int_0^t \theta(q_s - 1)^+\, ds - \int_0^t \mu\,(q_s \wedge 1)\, ds, \tag{4.1}$$

$$X_t = X_0 + \int_0^t \big(\alpha - \gamma(q_s - 1)^+ - \beta(q_s \wedge 1)\big)\, ds - \int_0^t \theta\big(\mathbf{1}_{\{q_s>1\}}\, X_s + \mathbf{1}_{\{q_s=1\}}\, X^+_s\big)\, ds$$
$$- \int_0^t \mu\big(\mathbf{1}_{\{q_s<1\}}\, X_s + \mathbf{1}_{\{q_s=1\}}\, X_s \wedge 0\big)\, ds + \int_0^t \sqrt{\lambda + \theta(q_s - 1)^+ + \mu\,(q_s \wedge 1)}\, dW_s\,. \tag{4.2}$$

First, I investigate stationary solutions of (4.1).

Lemma 4 *If* $\lambda \ge \mu$, *then* $\lim_{t\to\infty} q_t = (\lambda - \mu)/\theta + 1$. *If* $\lambda \le \mu$, *then* $\lim_{t\to\infty} q_t = \lambda/\mu$. *For all* t, $q_t \ne 1$ *except when* $q_0 = 1$ *and* $\lambda = \mu$ *in which case* $q_t = 1$ *for all* t *.*

Proof Suppose that $\lambda > \mu$. Then by (4.1),

$$\frac{d}{dt}\bigl(q_t - \frac{\lambda-\mu}{\theta} - 1\bigr)^2 = 2\bigl(q_t - \frac{\lambda-\mu}{\theta} - 1\bigr)(\lambda - \theta(q_t-1)^+ - \mu\,(q_t \wedge 1))\,.$$

If $q_t - (\lambda-\mu)/\theta - 1 \ge \epsilon$ for $\epsilon > 0$, then $\lambda - \theta(q_t-1)^+ - \mu\,(q_t \wedge 1) = \lambda - \theta(q_t-1) - \mu \le -\theta\epsilon$. If $q_t - (\lambda-\mu)/\theta - 1 \le -\epsilon$ for $\epsilon \in (0,(\lambda-\mu)/\theta)$, then $\lambda - \theta(q_t-1)^+ - \mu\,(q_t \wedge 1) \ge \lambda - \theta(q_t-1)^+ - \mu \ge \theta\epsilon$. Hence, $q_t \to (\lambda-\mu)/\theta + 1$ as $t \to \infty$.

If $\lambda < \mu$, then a similar reasoning applied to the function $(q_t - \lambda/\mu)^2$ shows that $q_t \to \lambda/\mu$. Suppose $\lambda = \mu$. Then $\dot q_t = \mu(1-q_t)^+ - \theta(q_t-1)^+$. Hence, $(d/dt)(q_t-1)^2 = 2(q_t-1)\bigl(\mu(1-q_t)^+ - \theta(q_t-1)^+\bigr) \le -2(\mu\wedge\theta)(q_t-1)^2$. Consequently, $q_t \to 1$.

The Markov chain Q^n is a birth-and-death process on $\mathbb{Z}_+$ with birth rates λ^n and death rates $\mu^n(i \wedge n) + \theta^n(i-n)^+$. Since $\sum_{k=1}^\infty (\lambda^n)^k / \prod_{i=1}^k \bigl(\mu^n(i\wedge n) + \theta^n(i-n)^+\bigr) < \infty$, it admits a unique stationary distribution which is a limit in the distance of total variation of the transient distributions for any initial condition. Let $\hat Q^n = (\hat Q^n_t,\, t \in \mathbb{R}_+)$ represent the stationary version of Q^n and let $\hat q_0 = \lim_{t\to\infty} q_t$.

Theorem 5 *Suppose that $\lambda^n/n \to \lambda$, that $\mu^n \to \mu$, and that $\theta^n \to \theta$ as $n \to \infty$. Then, for all $\epsilon > 0$ and $L > 0$,*

$$\lim_{n\to\infty} \mathbf{P}\bigl(\sup_{t\in[0,L]} \bigl|\frac{\hat Q^n_t}{n} - \hat q_0\bigr| > \epsilon\bigr) = 0\,.$$

Proof By part 2(b) of Lemma 6, $\sup_{n\in\mathbb{N}} \mathbf{E}\hat Q^n_0/n < \infty$, so the sequence $\hat Q^n_0/n$ is tight. By Theorem 1, the sequence of processes $(\hat Q^n_t/n,\, t\in\mathbb{R}_+)$ is tight and any limit point $(q_t,\, t \in \mathbb{R}_+)$ for convergence in distribution is a solution to (4.1) for a suitable $\mathbb{R}_+$-valued random variable q_0 where, by Fatou's lemma, $\mathbf{E}q_0 < \infty$. Since $(\hat Q^n_t/n,\, t\in\mathbb{R}_+)$ is stationary, so is $(q_t,\, t\in\mathbb{R}_+)$. By the proof of Lemma 4, $|q_t - \hat q_0|$ decreases in t and tends to zero as $t \to \infty$, so by dominated convergence $\mathbf{E}|q_t - \hat q_0| \to 0$. By stationarity, $\mathbf{E}|q_t - \hat q_0| = 0$.

Let process $\hat X^n = (\hat X^n(t),\, t \in \mathbb{R}_+)$ represent the stationary version of X^n ,i.e., $\hat X^n(t) = \sqrt{n}(\hat Q^n_t/n - \hat q_0)$.

Theorem 6 *Suppose that $\sqrt{n}(\lambda^n/n - \lambda) \to \alpha$, $\sqrt{n}(\mu^n - \mu) \to \beta$, and $\sqrt{n}(\theta^n - \theta) \to \gamma$ as $n \to \infty$. Then the processes $\hat X^n$ converge in distribution in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ as $n\to\infty$ to a stationary continuous-path Markov process $\hat X = (\hat X_t,\, t\in\mathbb{R}_+)$.*

If $\lambda < \mu$, then the process $\hat X$ is Gaussian with $\mathbf{E}\hat X_t = \alpha/\mu - \beta\,\lambda/\mu^2$ and $\mathbf{Cov}\,(\hat X_u, \hat X_v) = (\lambda/\mu)e^{-\mu|u-v|}$. If $\lambda > \mu$, then the process $\hat X$ is Gaussian with $\mathbf{E}\hat X_t = \alpha/\theta - \gamma(\lambda-\mu)/\theta^2 - \beta/\theta$ and $\mathbf{Cov}\,(\hat X_u,\hat X_v) = (\lambda/\theta)e^{-\theta|u-v|}$. If $\lambda = \mu$, then

$$\hat X_t = \hat X_0 + (\alpha-\beta)t - \theta\int_0^t \hat X_s^+\,ds + \mu\int_0^t (-\hat X_s)^+\,ds + \sqrt{2\mu}\hat W_t\,,$$

where the distribution of $\hat X_0$ has density $C\exp\bigl(\bigl((\alpha-\beta)x - (x^2/2)(\theta\,\mathbf{1}_{\{x\ge 0\}} + \mu\,\mathbf{1}_{\{x<0\}}\,)\bigr)/\mu\bigr)$, $(\hat W_t,\, t\in\mathbb{R}_+)$ is a standard Wiener process, and $\hat X_0$ and $(\hat W_t,\, t\in\mathbb{R}_+)$ are independent.

Proof The distributions of the random variables X^n_t with $Q^n_0 = q_0 = 0$ converge in the distance of total variation as $t\to\infty$ to the distribution of $\hat X^n_0$. By part 2(c) of Lemma 6 and Fatou's lemma, $\sup_{n\in\mathbb{N}} \mathbf{E}(\hat X^n_0)^2 < \infty$, so the sequence of the distributions of the $\hat X^n_0$ is tight. By Theorem 2, the sequence of the distributions of the $\hat X^n$ is tight and any limit point in distribution $(\grave X_t,\, t\in\mathbb{R}_+)$ satisfies the equation

$$\grave X_t = \grave X_0 + \bigl(\alpha - \gamma(\hat q_0 - 1)^+ - \beta(\hat q_0 \wedge 1)\bigr)\,t - \int_0^t \theta\bigl(\,\mathbf{1}_{\{\hat q_0 > 1\}}\,\grave X_s + \mathbf{1}_{\{\hat q_0 = 1\}}\,\grave X_s^+\bigr)\,ds$$
$$- \int_0^t \mu\bigl(\,\mathbf{1}_{\{\hat q_0 < 1\}}\,\grave X_s + \mathbf{1}_{\{\hat q_0 = 1\}}\,(\grave X_s \wedge 0)\bigr)\,ds + \sqrt{\lambda + \theta(\hat q_0 - 1)^+ + \mu\,(\hat q_0 \wedge 1)}\,\grave W_t,$$

where $(\grave W_t,\, t\in\mathbb{R}_+)$ is a standard Wiener process and $\grave X_0$ and $(\grave W_t,\, t\in\mathbb{R}_+)$ are independent. Since the X^n are stationary, so is $\grave X$. Stationary distributions of one-dimensional diffusions are available in the literature, see, e.g., Skorokhod [14].

Remark 5 Fleming, Simon, and Stolyar [3] obtain the distribution of $\hat{X}_t$ provided $\lambda = \mu$ starting with an explicit formula for the stationary distribution of Q^n .

Remark 6 If $\lambda \le \mu$, then the condition that $\sqrt{n}(\theta^n - \theta) \to \gamma$ can be disposed of and one can merely require that $\theta^n \to \theta$, as in Theorem 2 in Garnett, Mandelbaum, and Reiman [4].

Remark 7 The limits obtained in Theorems 2.1 and 2.3 in Whitt [17] correspond to the case where $\lambda^n = n\lambda$, $\mu^n = \mu$, $\theta^n = \theta$, and $\lambda > \mu$, so $\alpha = \beta = \gamma = 0$.

A Appendix

Lemma 5 *Let $(F(t),\, t \in \mathbb{R}_+)$ be a function of locally bounded variation and $(f(t),\, t \in \mathbb{R}_+)$ be a locally bounded Lebesgue measurable function. If a locally integrable function $(y(t),\, t \in \mathbb{R}_+)$ is such that $y(t) \prec F(t) - \int_0^t f(s)y(s)\, ds$, then*

$$y(t) \le e^{-\int_0^t f(s)\, ds} F(0) + e^{-\int_0^t f(s)\, ds} \int\limits_0^t e^{\int_0^s f(u)\, du}\, dF(s)\,.$$

Proof Let $g(t) = F(t) - \int_0^t f(s)y(s)\, ds - y(t)$. The function $(g(t))$ is nondecreasing, $g(0) \ge 0$, and

$$y(t) = F(t) - g(t) - \int\limits_0^t f(s)y(s)\, ds\,.$$

Hence,

$$\begin{aligned} y(t) &= e^{-\int_0^t f(s)\, ds}(F(0) - g(0)) + e^{-\int_0^t f(s)\, ds} \int\limits_0^t e^{\int_0^s f(u)\, du}\, d(F(s) - g(s)) \\ &\le e^{-\int_0^t f(s)\, ds} F(0) + e^{-\int_0^t f(s)\, ds} \int\limits_0^t e^{\int_0^s f(u)\, du}\, dF(s)\,. \end{aligned}$$

The next lemma provides the bounds that have been used for the analysis of large-time behaviour. Let $T > 0$ and $\sigma_s^n = |\alpha_s^n - \gamma_s^n(q_s - \kappa_s)^+ - \beta_s^n(q_s \wedge \kappa_s)| + |(\theta_s^n - \mu_s^n)\delta_s^n|$. Let me recall that q_t is defined by (2.7).

Lemma 6 *1. (a) If the functions $(\lambda_t,\, t \in \mathbb{R}_+)$, $(\mu_t,\, t \in \mathbb{R}_+)$, and $(\theta_t,\, t \in \mathbb{R}_+)$ are T-periodic and $\int_0^T (\mu_s \wedge \theta_s)\, ds > 0$, then, for all $t \in \mathbb{R}_+$,*

$$q_t \le e^{-\lfloor t/T \rfloor \int_0^T (\mu_s \wedge \theta_s)\, ds} q_0 + \frac{e^{\int_0^T (\mu_s \wedge \theta_s)\, ds}}{1 - e^{-\int_0^T (\mu_s \wedge \theta_s)\, ds}} \int\limits_0^T \lambda_s\, ds\,.$$

(b) If the functions $(\lambda_t^n,\, t \in \mathbb{R}_+)$, $(\mu_t^n,\, t \in \mathbb{R}_+)$, and $(\theta_t^n,\, t \in \mathbb{R}_+)$ are T-periodic and $\int_0^T (\mu_s^n \wedge \theta_s^n)\, ds > 0$, then, for all $t \in \mathbb{R}_+$ and $V > 0$,

$$\mathbf{E} Q_t^n\, \mathbf{1}_{\{Q_0^n \le V\}} \le e^{-\lfloor t/T \rfloor \int_0^T (\mu_s^n \wedge \theta_s^n)\, ds}\, \mathbf{E} Q_0^n\, \mathbf{1}_{\{Q_0^n \le V\}} + \frac{e^{\int_0^T (\mu_s^n \wedge \theta_s^n)\, ds}}{1 - e^{-\int_0^T (\mu_s^n \wedge \theta_s^n)\, ds}} \int\limits_0^T \lambda_s^n\, ds\,.$$

.

(c) If the functions $(\lambda_t^n,\, t \in \mathbb{R}_+)$, $(\mu_t^n,\, t \in \mathbb{R}_+)$, $(\theta_t^n,\, t \in \mathbb{R}_+)$, $(\lambda_t,\, t \in \mathbb{R}_+)$, $(\mu_t,\, t \in \mathbb{R}_+)$, and $(\theta_t,\, t \in \mathbb{R}_+)$ are T-periodic, $\int_0^T (\mu_s \wedge \theta_s)\, ds > 0$, and, for some $\epsilon > 0$, $\int_0^T \big(2(\mu_s^n \wedge \theta_s^n) - \epsilon\, \sigma_s^n - (\mu_s^n \vee \theta_s^n)/(2\sqrt{n})\big)\, ds > 0$, then, for all $t \in \mathbb{R}_+$ and $V > 0$,

$$\begin{aligned} &\mathbf{E}(X_t^n)^2\, \mathbf{1}_{\{|X_0^n| \le V\}} \\ &\quad \le e^{-\lfloor t/T \rfloor \int_0^T \big(2(\mu_s^n \wedge \theta_s^n) - \epsilon\, \sigma_s^n - (\mu_s^n \vee \theta_s^n)/(2\sqrt{n})\big)\, ds + \int_0^T |2(\mu_s^n \wedge \theta_s^n) - \epsilon\, \sigma_s^n - (\mu_s^n \vee \theta_s^n)/(2\sqrt{n})|\, ds}\, \mathbf{E}(X_0^n)^2\, \mathbf{1}_{\{|X_0^n| \le V\}} \\ &\quad + \frac{e^{2\int_0^T |2(\mu_s^n \wedge \theta_s^n) - \epsilon\, \sigma_s^n - (\mu_s^n \vee \theta_s^n)/(2\sqrt{n})|\, ds}}{1 - e^{-\int_0^T \big(2(\mu_s^n \wedge \theta_s^n) - \epsilon\, \sigma_s^n - (\mu_s^n \vee \theta_s^n)/(2\sqrt{n})\big)\, ds}} \int\limits_0^T \Big(\frac{1}{\epsilon}\, \sigma_s^n + \frac{\lambda_s^n}{n} \\ &\qquad + (\mu_s^n \vee \theta_s^n) \sup_{u \in \mathbb{R}_+} q_u + \frac{1}{2\sqrt{n}} (\mu_s^n \vee \theta_s^n)\Big)\, ds\,. \end{aligned}$$

2. (a) If $\sup_{t\in\mathbb{R}_+}\lambda_t<\infty$ *and* $\inf_{t\in\mathbb{R}_+}(\mu_t\wedge\theta_t)>0$*, then, for all* $t\in\mathbb{R}_+$*,*

$$q_t\le e^{-\int_0^t(\mu_s\wedge\theta_s)\,ds}q_0+\frac{\sup_{s\in\mathbb{R}_+}\lambda_s}{\inf_{s\in\mathbb{R}_+}(\theta_s\wedge\mu_s)}\,.$$

(b) If $\sup_{t\in\mathbb{R}_+}\lambda_t^n<\infty$ *and* $\inf_{t\in\mathbb{R}_+}(\mu_t^n\wedge\theta_t^n)>0$*, then, for all* $t\in\mathbb{R}_+$ *and* $V>0$*,*

$$\mathbf{E}Q_t^n\,\mathbf{1}_{\{Q_0^n\le V\}}\le e^{-\int_0^t(\mu_s^n\wedge\theta_s^n)\,ds}\,\mathbf{E}Q_0^n\,\mathbf{1}_{\{Q_0^n\le V\}}+\frac{\sup_{s\in\mathbb{R}_+}\lambda_s^n}{\inf_{s\in\mathbb{R}_+}(\theta_s^n\wedge\mu_s^n)}\,.$$

(c) If $\sup_{t\in\mathbb{R}_+}\lambda_t<\infty$, $\inf_{t\in\mathbb{R}_+}(\mu_t\wedge\theta_t)>0$, $\sup_{t\in\mathbb{R}_+}\lambda_t^n<\infty$, $\sup_{t\in\mathbb{R}_+}\theta_t^n<\infty$, $\sup_{t\in\mathbb{R}_+}\mu_t^n<\infty$, $\sup_{t\in\mathbb{R}_+}\sigma_t^n<\infty$*, and* $\inf_{t\in\mathbb{R}_+}\big(2(\mu_t^n\wedge\theta_t^n)-\epsilon\,\sigma_t^n-(\mu_t^n\vee\theta_t^n)/(2\sqrt{n})\big)>0$ *for some* $\epsilon>0$*, then, for all* $t\in\mathbb{R}_+$ *and* $V>0$*,*

$$\mathbf{E}(X_t^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}}\le e^{-\int_0^t\big(2(\mu_s^n\wedge\theta_s^n)-\epsilon\,\sigma_s^n-(\mu_s^n\vee\theta_s^n)/(2\sqrt{n})\big)\,ds}\,\mathbf{E}(X_0^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}}+\frac{\sup_{s\in\mathbb{R}_+}\big(\sigma_s^n/\epsilon+\lambda_s^n/n+(\mu_s^n\vee\theta_s^n)q_s+(\mu_s^n\vee\theta_s^n)/(2\sqrt{n})\big)}{\inf_{s\in\mathbb{R}_+}\big(2(\mu_s^n\wedge\theta_s^n)-\epsilon\,\sigma_s^n-(\mu_s^n\vee\theta_s^n)/(2\sqrt{n})\big)}\,.$$

Proof Let me start with part 1(b). Since $Q_t^n\le Q_0^n+A_t^n$ by (2.1), I have that $\mathbf{E}Q_t^n\,\mathbf{1}_{\{Q_0^n\le V\}}\le V+\int_0^t\lambda_s^n\,ds$. By Lemma 1, the processes $(M_t^{n,A}\,\mathbf{1}_{\{Q_0^n\le V\}}\,,\,t\in\mathbb{R}_+)$, $(M_t^{n,R}\,\mathbf{1}_{\{Q_0^n\le V\}}\,,\,t\in\mathbb{R}_+)$, and $(M_t^{n,B}\,\mathbf{1}_{\{Q_0^n\le V\}}\,,\,t\in\mathbb{R}_+)$ are $\mathbf{F}^n$-locally square integrable martingales with respective predictable quadratic variation processes $(\,\mathbf{1}_{\{Q_0^n\le V\}}\int_0^t\lambda_s^n\,ds\,,t\in\mathbb{R}_+)$, $(\,\mathbf{1}_{\{Q_0^n\le V\}}\int_0^t\theta_s^n\,(Q_s^n-K_s^n)^+\,ds,\,t\in\mathbb{R}_+)$, and $(\,\mathbf{1}_{\{Q_0^n\le V\}}\int_0^t\mu_s^n\,(Q_s^n\wedge K_s^n)\,ds,\,t\in\mathbb{R}_+)$. Since the latter processes are of finite expectation, I obtain that $\mathbf{E}(M_t^{n,A})^2\,\mathbf{1}_{\{Q_0^n\le V\}}<\infty$, $\mathbf{E}(M_t^{n,R})^2\,\mathbf{1}_{\{Q_0^n\le V\}}<\infty$, and $\mathbf{E}(M_t^{n,B})^2\,\mathbf{1}_{\{Q_0^n\le V\}}<\infty$. In particular, $(M_t^{n,A}\,\mathbf{1}_{\{Q_0^n\le V\}}\,,\,t\in\mathbb{R}_+)$, $(M_t^{n,R}\,\mathbf{1}_{\{Q_0^n\le V\}}\,,\,t\in\mathbb{R}_+)$, and $(M_t^{n,B}\,\mathbf{1}_{\{Q_0^n\le V\}}\,,\,t\in\mathbb{R}_+)$ are martingales. By (2.3),

$$\mathbf{E}Q_t^n\,\mathbf{1}_{\{Q_0^n\le V\}}\prec\mathbf{E}Q_0^n\,\mathbf{1}_{\{Q_0^n\le V\}}+\int_0^t\lambda_s^n\,ds-\int_0^t(\mu_s^n\wedge\theta_s^n)\,\mathbf{E}Q_s^n\,\mathbf{1}_{\{Q_0^n\le V\}}\,ds\,. \tag{A.1}$$

By Lemma 5,

$$\mathbf{E}Q_t^n\,\mathbf{1}_{\{Q_0^n\le V\}}\le e^{-\int_0^t(\mu_s^n\wedge\theta_s^n)\,ds}\mathbf{E}Q_0^n\,\mathbf{1}_{\{Q_0^n\le V\}}+e^{-\int_0^t(\mu_s^n\wedge\theta_s^n)\,ds}\int_0^t e^{\int_0^s(\mu_u^n\wedge\theta_u^n)\,du}\lambda_s^n\,ds\,. \tag{A.2}$$

If $vT\le t<(v+1)T$, where $v\in\mathbb{Z}_+$, then by T-periodicity,

$$\begin{aligned}
e^{-\int_0^t(\mu_s^n\wedge\theta_s^n)\,ds}\int_0^t e^{\int_0^s(\mu_u^n\wedge\theta_u^n)\,du}\lambda_s^n\,ds&\le e^{-\int_0^{vT}(\mu_s^n\wedge\theta_s^n)\,ds}\sum_{i=1}^{v+1}e^{\int_0^{iT}(\mu_u^n\wedge\theta_u^n)\,du}\int_{(i-1)T}^{iT}\lambda_s^n\,ds\\
&=e^{-v\int_0^T(\mu_s^n\wedge\theta_s^n)\,ds}\sum_{i=1}^{v+1}e^{i\int_0^T(\mu_u^n\wedge\theta_u^n)\,du}\int_0^T\lambda_s^n\,ds=e^{-(v-1)\int_0^T(\mu_s^n\wedge\theta_s^n)\,ds}\frac{e^{(v+1)\int_0^T(\mu_u^n\wedge\theta_u^n)\,du}-1}{e^{\int_0^T(\mu_u^n\wedge\theta_u^n)\,du}-1}\int_0^T\lambda_s^n\,ds\\
&\le\frac{e^{2\int_0^T(\mu_u^n\wedge\theta_u^n)\,du}}{e^{\int_0^T(\mu_u^n\wedge\theta_u^n)\,du}-1}\int_0^T\lambda_s^n\,ds\,.
\end{aligned}$$

In addition, $e^{-\int_0^t(\mu_s^n\wedge\theta_s^n)\,ds}\le e^{-\lfloor t/T\rfloor\int_0^T(\mu_s^n\wedge\theta_s^n)\,ds}$. Part 1(b) has been proved.

Part 1(a) follows by a similar argument if one observes that by (2.7),

$$q_t\prec q_0+\int_0^t\lambda_s\,ds-\int_0^t(\theta_s\wedge\mu_s)q_s\,ds$$

so that, by Lemma 5,

$$q_t\le e^{-\int_0^t(\theta_s\wedge\mu_s)\,ds}q_0+e^{-\int_0^t(\theta_s\wedge\mu_s)\,ds}\int_0^t e^{\int_0^s(\theta_u\wedge\mu_u)\,du}\lambda_s\,ds\,. \tag{A.3}$$

In order to prove part 1(c), let me note that by (2.14),

$$
\begin{aligned}
(X_t^n)^2 &= (X_0^n)^2 + 2\int_0^t X_{s-}^n\,dX_s^n + \sum_{0<s\le t}(\Delta X_s^n)^2 = (X_0^n)^2 + 2\int_0^t \alpha_s^n X_s^n\,ds \\
&\quad - 2\int_0^t \theta_s^n X_s^n\Big(\big(X_s^n - \delta_s^n + \sqrt{n}(q_s-\kappa_s)\big)^+ - \sqrt{n}(q_s-\kappa_s)^+\Big)\,ds - 2\int_0^t \gamma_s^n (q_s-\kappa_s)^+ X_s^n\,ds \\
&\quad - 2\int_0^t \mu_s^n X_s^n\Big(\big(X_s^n + \sqrt{n}q_s\big)\wedge(\delta_s^n + \sqrt{n}\kappa_s) - \sqrt{n}\,(q_s\wedge\kappa_s)\Big)\,ds - 2\int_0^t \beta_s^n(q_s\wedge\kappa_s)X_s^n\,ds \\
&\quad + \frac{2}{\sqrt{n}}\int_0^t X_{s-}^n\,dM_s^{n,A} - \frac{2}{\sqrt{n}}\int_0^t X_{s-}^n\,dM_s^{n,R} - \frac{2}{\sqrt{n}}\int_0^t X_{s-}^n\,dM_s^{n,B} \\
&\quad + \frac{1}{n}\sum_{0<s\le t}(\Delta M_s^{n,A})^2 + \frac{1}{n}\sum_{0<s\le t}(\Delta M_s^{n,R})^2 + \frac{1}{n}\sum_{0<s\le t}(\Delta M_s^{n,B})^2\,.
\end{aligned}
$$

On noting that

$$
\begin{aligned}
&\theta_s^n X_s^n\Big(\big(X_s^n - \delta_s^n + \sqrt{n}(q_s-\kappa_s)\big)^+ - \sqrt{n}(q_s-\kappa_s)^+\Big) + \mu_s^n X_s^n\Big(\big(X_s^n + \sqrt{n}q_s\big)\wedge(\delta_s^n + \sqrt{n}\kappa_s) - \sqrt{n}\,(q_s\wedge\kappa_s)\Big) \\
&\qquad \ge (\mu_s^n\wedge\theta_s^n)(X_s^n)^2 - |(\theta_s^n - \mu_s^n)\delta_s^n X_s^n|\,,
\end{aligned}
$$

I obtain that, for $\epsilon>0$,

$$
\begin{aligned}
(X_t^n)^2 &\prec (X_0^n)^2 + 2\int_0^t \sigma_s^n|X_s^n|\,ds - 2\int_0^t(\mu_s^n\wedge\theta_s^n)(X_s^n)^2\,ds + \frac{2}{\sqrt{n}}\int_0^t X_{s-}^n\,dM_s^{n,A} - \frac{2}{\sqrt{n}}\int_0^t X_{s-}^n\,dM_s^{n,R} \\
&\quad - \frac{2}{\sqrt{n}}\int_0^t X_{s-}^n\,dM_s^{n,B} + \frac{1}{n}\sum_{0<s\le t}(\Delta M_s^{n,A})^2 + \frac{1}{n}\sum_{0<s\le t}(\Delta M_s^{n,R})^2 + \frac{1}{n}\sum_{0<s\le t}(\Delta M_s^{n,B})^2 \\
&\prec (X_0^n)^2 + \frac{1}{\epsilon}\int_0^t \sigma_s^n\,ds - \int_0^t\big(2(\mu_s^n\wedge\theta_s^n) - \epsilon\sigma_s^n\big)(X_s^n)^2\,ds + \frac{2}{\sqrt{n}}\int_0^t X_{s-}^n\,dM_s^{n,A} - \frac{2}{\sqrt{n}}\int_0^t X_{s-}^n\,dM_s^{n,R} \\
&\quad - \frac{2}{\sqrt{n}}\int_0^t X_{s-}^n\,dM_s^{n,B} + \frac{1}{n}\sum_{0<s\le t}(\Delta M_s^{n,A})^2 + \frac{1}{n}\sum_{0<s\le t}(\Delta M_s^{n,R})^2 + \frac{1}{n}\sum_{0<s\le t}(\Delta M_s^{n,B})^2\,.
\end{aligned}
$$

Hence, for $V>0$,

$$
\begin{aligned}
(X_t^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}} &\prec (X_0^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}} + \frac{1}{\epsilon}\int_0^t \sigma_s^n\,\mathbf{1}_{\{|X_0^n|\le V\}}\,ds - \int_0^t\big(2(\mu_s^n\wedge\theta_s^n) - \epsilon\,\sigma_s^n\big)(X_s^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}}\,ds \\
&\quad + \frac{2}{\sqrt{n}}\int_0^t X_{s-}^n\,dM_s^{n,A,V} - \frac{2}{\sqrt{n}}\int_0^t X_{s-}^n\,dM_s^{n,R,V} - \frac{2}{\sqrt{n}}\int_0^t X_{s-}^n\,dM_s^{n,B,V} \\
&\quad + \frac{1}{n}\sum_{0<s\le t}(\Delta M_s^{n,A,V})^2 + \frac{1}{n}\sum_{0<s\le t}(\Delta M_s^{n,R,V})^2 + \frac{1}{n}\sum_{0<s\le t}(\Delta M_s^{n,B,V})^2 \qquad \text{(A.4)}
\end{aligned}
$$

where $M_s^{n,i,V} = M_s^{n,i}\,\mathbf{1}_{\{|X_0^n|\le V\}}$, for $i=A,R,B$.

By Lemma 1, the processes $M^{n,i,V} = (M_t^{n,i,V},\,t\in\mathbb{R}_+)$ are $\mathbf{F}^n$-locally square integrable martingales with predictable quadratic variation processes $\langle M^{n,i,V}\rangle = \langle M^{n,i}\rangle\,\mathbf{1}_{\{|X_0^n|\le V\}}$. Since $Q_t^n \le Q_0^n + A_t^n$ by (2.1) and $\mathbf{E}(A_t^n)^2 = \int_0^t \lambda_s^n\,ds + \big(\int_0^t \lambda_s^n\,ds\big)^2 < \infty$, I have that $\mathbf{E}(Q_t^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}} < \infty$. Hence, by (2.4a), (2.4b), (2.4c), $\mathbf{E}\langle M^{n,i,V}\rangle_t < \infty$, which implies that $\mathbf{E}\big(\sup_{s\le t}(M_s^{n,i,V})^2\big) < \infty$, that the $M^{n,i,V}$ are $\mathbf{F}^n$-martingales, and that $\mathbf{E}(M_t^{n,i,V})^2 = \mathbf{E}\langle M^{n,i,V}\rangle_t$. Consequently, the processes $\big(\int_0^t X_{s-}^n\,dM_s^{n,i,V},\,t\in\mathbb{R}_+\big)$ are $\mathbf{F}^n$-martingales. Since the $M^{n,i,V}$ are purely discontinuous locally square integrable martingales by being of locally bounded variation, $\mathbf{E}\sum_{0<s\le t}(\Delta M_s^{n,i,V})^2 = \mathbf{E}\langle M^{n,i,V}\rangle_t$.

On taking expectations in (A.4),

$$\mathbf{E}(X_t^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}}\prec\mathbf{E}(X_0^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}}+\frac{1}{\epsilon}\int_0^t\sigma_s^n\,ds+\frac{1}{n}\big(\mathbf{E}\langle M^{n,A}\rangle_t\,\mathbf{1}_{\{|X_0^n|\le V\}}+\mathbf{E}\langle M^{n,R}\rangle_t\,\mathbf{1}_{\{|X_0^n|\le V\}}+\mathbf{E}\langle M^{n,B}\rangle_t\,\mathbf{1}_{\{|X_0^n|\le V\}}\big)-\int_0^t\big(2(\mu_s^n\wedge\theta_s^n)-\epsilon\,\sigma_s^n\big)\mathbf{E}(X_s^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}}\,ds\,.$$

By (2.4a), (2.4b), (2.4c), and (2.13),

$$\frac{1}{n}\big(\mathbf{E}\langle M^{n,A}\rangle_t\,\mathbf{1}_{\{|X_0^n|\le V\}}+\mathbf{E}\langle M^{n,R}\rangle_t\,\mathbf{1}_{\{|X_0^n|\le V\}}+\mathbf{E}\langle M^{n,B}\rangle_t\,\mathbf{1}_{\{|X_0^n|\le V\}}\big)\prec\int_0^t\frac{\lambda_s^n}{n}\,ds+\int_0^t(\mu_s^n\vee\theta_s^n)\frac{\mathbf{E}Q_s^n\,\mathbf{1}_{\{|X_0^n|\le V\}}}{n}ds\prec\int_0^t\frac{\lambda_s^n}{n}\,ds+\int_0^t(\mu_s^n\vee\theta_s^n)\big(\frac{\mathbf{E}X_s^n\,\mathbf{1}_{\{|X_0^n|\le V\}}}{\sqrt{n}}+q_s\big)\,ds$$
$$\prec\int_0^t\frac{\lambda_s^n}{n}\,ds+\int_0^t(\mu_s^n\vee\theta_s^n)\big(q_s+\frac{1}{2\sqrt{n}}\big)\,ds+\frac{1}{2\sqrt{n}}\int_0^t(\mu_s^n\vee\theta_s^n)\,\mathbf{E}(X_s^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}}\,ds\,.$$

Thus, for $t\in\mathbb{R}_+$,

$$\mathbf{E}(X_t^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}}\prec\mathbf{E}(X_0^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}}+\frac{1}{\epsilon}\int_0^t\sigma_s^n\,ds+\int_0^t\frac{\lambda_s^n}{n}\,ds+\int_0^t(\mu_s^n\vee\theta_s^n)\big(q_s+\frac{1}{2\sqrt{n}}\big)\,ds-\int_0^t\big(2(\mu_s^n\wedge\theta_s^n)-\epsilon\,\sigma_s^n-\frac{1}{2\sqrt{n}}(\mu_s^n\vee\theta_s^n)\big)\mathbf{E}(X_s^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}}\,ds\,.$$

By Lemma 5,

$$\mathbf{E}(X_t^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}}\le e^{-\int_0^t\big(2(\mu_s^n\wedge\theta_s^n)-\epsilon\,\sigma_s^n-(\mu_s^n\vee\theta_s^n)/(2\sqrt{n})\big)\,ds}\mathbf{E}(X_0^n)^2\,\mathbf{1}_{\{|X_0^n|\le V\}}+e^{-\int_0^t\big(2(\mu_s^n\wedge\theta_s^n)-\epsilon\,\sigma_s^n-(\mu_s^n\vee\theta_s^n)/(2\sqrt{n})\big)\,ds}\int_0^t e^{\int_0^s\big(2(\mu_u^n\wedge\theta_u^n)-\epsilon\,\sigma_u^n-(\mu_u^n\vee\theta_u^n)/(2\sqrt{n})\big)\,du}\big(\frac{\sigma_s^n}{\epsilon}+\frac{\lambda_s^n}{n}+(\mu_s^n\vee\theta_s^n)\big(q_s+\frac{1}{2\sqrt{n}}\big)\big)\,ds\,. \tag{A.5}$$

In analogy with the earlier argument, if $vT\le t<(v+1)T$, where $v\in\mathbb{Z}_+$, recalling that $\sup_{u\in\mathbb{R}_+}q_u<\infty$ by part 1(a),

$$e^{-\int_0^t\big(2(\mu_s^n\wedge\theta_s^n)-\epsilon\,\sigma_s^n-(\mu_s^n\vee\theta_s^n)/(2\sqrt{n})\big)\,ds}\int_0^t e^{\int_0^s\big(2(\mu_u^n\wedge\theta_u^n)-\epsilon\,\sigma_u^n-(\mu_u^n\vee\theta_u^n)/(2\sqrt{n})\big)\,du}\big(\frac{\sigma_s^n}{\epsilon}+\frac{\lambda_s^n}{n}+(\mu_s^n\vee\theta_s^n)\big(q_s+\frac{1}{2\sqrt{n}}\big)\big)\,ds\le e^{-v\int_0^T\big(2(\mu_s^n\wedge\theta_s^n)-\epsilon\,\sigma_s^n-(\mu_s^n\vee\theta_s^n)/(2\sqrt{n})\big)\,ds+2\int_0^T|2(\mu_s^n\wedge\theta_s^n)-\epsilon\,\sigma_s^n-(\mu_s^n\vee\theta_s^n)/(2\sqrt{n})|\,ds}$$
$$\sum_{i=0}^v e^{i\int_0^T\big(2(\mu_u^n\wedge\theta_u^n)-\epsilon\,\sigma_u^n-(\mu_u^n\vee\theta_u^n)/(2\sqrt{n})\big)\,du}\int_0^T\big(\frac{1}{\epsilon}\,\sigma_s^n+\frac{\lambda_s^n}{n}+(\mu_s^n\vee\theta_s^n)\big(\sup_{u\in\mathbb{R}_+}q_u+\frac{1}{2\sqrt{n}}\big)\big)\,ds$$
$$\le\frac{e^{2\int_0^T|2(\mu_u^n\wedge\theta_u^n)-\epsilon\,\sigma_u^n-(\mu_u^n\vee\theta_u^n)/(2\sqrt{n})|\,du}}{1-e^{-\int_0^T\big(2(\mu_u^n\wedge\theta_u^n)-\epsilon\,\sigma_u^n-(\mu_u^n\vee\theta_u^n)/(2\sqrt{n})\big)\,du}}\int_0^T\big(\frac{1}{\epsilon}\,\sigma_s^n+\frac{\lambda_s^n}{n}+(\mu_s^n\vee\theta_s^n)\big(\sup_{u\in\mathbb{R}_+}q_u+\frac{1}{2\sqrt{n}}\big)\big)\,ds\,,$$

where the last inequality uses the fact that $\int_0^T\big(2(\mu_u^n\wedge\theta_u^n)-\epsilon\,\sigma_u^n-(\mu_u^n\vee\theta_u^n)/(2\sqrt{n})\big)\,du>0$. The latter expression furnishes the required bound. Part 1 has been proved.

The assertions of part 2 also follow from the respective inequalities (A.2), (A.3), and (A.5). For instance part 2(b) is obtained by applying the bound

$$\int_0^t e^{\int_0^s(\theta_u^n\wedge\mu_u^n)\,du}\,\lambda_s^n\,ds\le\frac{\sup_{t\in\mathbb{R}_+}\lambda_t^n}{\inf_{t\in\mathbb{R}_+}(\theta_t^n\wedge\mu_t^n)}\,\big(e^{\int_0^t(\theta_s^n\wedge\mu_s^n)\,ds}-1\big)\,.$$

Acknowledgements I am thankful to the referees for the careful reading of the manuscript and insightful comments.

References

1. Ethier, S.N., Kurtz, T.G.: Markov Processes. Characterization and Convergence. Wiley (1986)
2. Fayolle, G., Malyshev, V.A., Men′shikov, M.V.: Topics in the constructive theory of countable Markov chains. Cambridge University Press, Cambridge (1995)
3. Fleming, P., Simon, B., Stolyar, A.: Heavy traffic limit for a mobile phone system loss model. In: Proc. 2nd Internat. Conf. Telecommunication Systems, Modeling, and Anal., pp. 158–176. Nashville, TN (1994)
4. Garnett, O., Mandelbaum, A., Reiman, M.: Designing a call center with impatient customers. Manufacturing Service Oper. Management **4**(3), 208–227 (2002)
5. Has'minskii, R.: Stochastic Stability of Differential Equations. Sijthoff & Noordhoff (1980). (Original title: Ustoicivost' sistem differencial'nyh uravnenii pri slucainyh vozmusceniyah ih parametrov, Nauka, Moscow, 1969)
6. Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes, 2nd edn. North Holland (1989)
7. Jacod, J., Shiryaev, A.: Limit Theorems for Stochastic Processes, *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol. 288, Springer-Verlag, Berlin (1987)
8. Liptser, R., Shiryayev, A.: Theory of Martingales. Kluwer (1989)
9. Mandelbaum, A., Massey, W., Reiman, M., Rider, B., Stolyar, A.: Queue lengths and waiting times for multiserver queues with abandonment and retrials. Telecommunication Systems **21**(2-4), 149–172, (2002)
10. Mandelbaum, A., Massey, W.A., Reiman, M.: Strong approximations for Markovian service networks. Queueing Syst. **30**, 149–201 (1998)
11. Meyn, S.P., Tweedie, R.L.: Markov chains and stochastic stability. Communications and Control Engineering Series. Springer-Verlag London Ltd., London (1993)
12. Pang, G., Talreja, R., Whitt, W.: Martingale proofs of many-server heavy-traffic limits for Markovian queues. Probab. Surv. **4**, 193 – 267 (electronic) (2007)
13. Puhalskii, A.A.: The $M_t/M_t/K_t + M_t$ queue in heavy traffic. Available at arXiv.org/abs/0807.4621
14. Skorokhod, A.V.: Asymptotic Methods in the Theory of Stochastic Differential Equations, *Translations of Mathematical Monographs*, vol. 78. American Mathematical Society, Providence, RI (1989)
15. Smorodinskiĭ, A.V.: Asymptotic distribution of queue length in a queueing system. Avtomat. i Telemekh. (2), 92–99 (1986)
16. Talreja, R., Whitt, W.: Heavy-traffic limits for waiting times in many-server queues with abandonment. Ann. Appl. Probab. **19**(6), 2137–2175 (2009). DOI 10.1214/09-AAP606. URL `http://0-dx.doi.org.skyline.ucdenver.edu/10.1214/09-AAP606`
17. Whitt, W.: Efficiency-driven heavy-traffic approximations for many-server queues with abandonment. Management Sci. **50**(10), 1449–1461 (2004)
18. Whitt, W.: Engineering solution of a basic call center model. Management Sci. **51**(2), 221–235 (2005)
19. Zeltyn, S., Mandelbaum, A.: Call centers with impatient customers: many-server asymptotics of the $M/M/n + G$ queue. Queueing Syst. **51**(3-4), 361–402 (2005)