Stochastic Systems
2017, Vol. 7, No. 1, 95–142
DOI: 10.1214/15-SSY175

HEAVY-TRAFFIC LIMIT FOR THE INITIAL CONTENT PROCESS

By A. Korhan Aras*, Yunan Liu* and Ward Whitt†

*North Carolina State University** and *Columbia University*†

To understand the performance of a queueing system, it can be useful to focus on the evolution of the content that is initially in service at some time. That necessarily will be the case in service systems that provide service during normal working hours each day, with the system shutting down at some time, except that all customers already in service at the termination time are allowed to complete their service. Also, for infinite-server queues, it is often fruitful to partition the content into the initial content and the new input, because these can be analyzed separately. With i.i.d service times having a non-exponential distribution, the state of the initial content can be described by specifying the elapsed service times of the remaining initial customers. That initial content process is then a Markov process. This paper establishes a many-server heavy-traffic (MSHT) functional central limit theorem (FCLT) for the initial content process in the space $\mathbb{D}_{\mathbb{D}}$, assuming a FCLT for the initial age process, with the number of customers initially in service growing in the limit. The proof applies a symmetrization lemma from the literature on empirical processes to address a technical challenge: For each time, including time 0, the conditional remaining service times, given the ages, are mutually independent but in general not identically distributed.

1. Introduction.

Heavy-traffic (HT) *functional central limit theorems* (FCLT's) for the standard $G/G/s$ queueing model, with unlimited waiting space and service in order of arrival, expose the impact of the stochastic variability in the arrival and service processes on the transient and steady-state performance. This is important because the general $G/G/s$ model is far less tractable than its Markovian $M/M/s$ counterpart, even for the special case in which the interarrival times and service times come from independent sequences of i.i.d. random variables. From [14, 41], we know that conventional heavy-traffic theory tells a simple story: With conventional heavy-traffic, where the arrival rate increases to the maximum possible service rate with a fixed number of servers, the arrival and service processes contribute via

Received February 2015.
AMS 2000 subject classifications: Primary 60K25; secondary 60F17, 90B22, 37C55
Keywords and phrases: Heavy-traffic limits for queues, many-server queues, infinite-server queues, terminating queues, time-varying arrivals, two-parameter processes, central limit theorem for non-identically distributed random variables.

their joint FCLT. Thus, with appropriate time and space scaling, we obtain the same reflected Brownian motion (RBM) limiting diffusion process for the $G/G/s$ model as for the $M/M/s$ model (and thus also for the $M/M/1$ model) except for a modification of the constant diffusion coefficient to account for the different variability. (For a discussion of interesting and important variability effects, see §9.6 of [41].) Moreover, assuming a deterministic limit for the scaled initial number in system, the initial conditions are not influenced by the variability at all, and only affect the initial state of the RBM.

Many-server heavy-traffic (MSHT) FCLT's tell a different story: With a MSHT FCLT, where both the arrival rate and number of servers increase without bound, and there is no extra time scaling, the variability in the service process and the initial conditions contribute in a more complicated way. Thus, the early MSHT FCLT in [13] was only for the $GI/M/s$ model, having i.i.d. exponential service times. With that condition, in Theorem 3 of [13] it was only necessary to assume that the properly scaled initial conditions converges to a nondegenerate limit. From §7.3 of [29], we know that MSHT limits for the $G/M/s$ model and the $G/M/s+M$ counterpart with customer abandonment from queue also depend on their general arrival processes only via its FCLT behavior, but non-M service processes influence the performance at all times through the service times that are in progress at those times.

Thus, in order to obtain a MSHT FCLT with a Markov limit process for models without i.i.d. exponential service times, except for very special cases [42], it is necessary to keep track of the elapsed service times and is convenient to greatly simplify the assumption for the initial conditions. For i.i.d. phase-type service-time distributions, we can keep track of the elapsed service times by keeping track of the number of phases of each type in service at each time, as in §4 of [13], Theorem 3 of [40], [33] and other papers. More generally, to obtain a Markov limit process, it is necessary to use two-parameter processes or measure-valued processes that keeps track of all the service times in process at each time as in [20, 17, 18, 19, 30, 31, 35].

However, even these new general results make strong simplifying assumptions about the initial conditions. That is so even in the *infinite-server* (IS) setting, where (under regularity conditions) the old content can be analyzed separately from the new content. For example, Pang and Whitt [30, 31] assumed that the system starts empty or with i.i.d. remaining service times.

Of course, if we are considering a stationary model, then we are usually interested in the steady-state distribution. Clearly, the steady-state distribution should be independent of the initial conditions under general regularity conditions. Thus, when we are interested in the steady-state distribution of a stationary model, there is little interest in the initial conditions; then it is of only minor technical interest to show that the steady-state distribution is independent of the initial conditions, in both pre-limit and limit processes.

In contrast, here we are motivated by the desire to develop asymptotic approximations for the time-varying behavior of nonstationary models, having time-varying arrival rates. However, even then, we may be unconcerned about initial conditions. Under regularity conditions, we anticipate that the time-varying behavior of a nonstationary models, having a time-varying arrival rate, will be independent of the initial conditions after a reasonable amount of time has passed. Concrete results in that direction are provided for the $G_t/M_t/s_t + GI_t$ fluid model in [22]. In particular, the existence of periodic limits and a more general asymptotic lack-of-memory property are established.

However, here we are primarily motivated by the desire to develop asymptotic approximations that apply to time-varying behavior of nonstationary models over shorter time intervals, where both the initial conditions and the new input may contribute significantly to system performance. We are also interested in describing system performance after an arrival process has been turned off. To the best of our knowledge, this is the first paper to address these problems.

In particular, in this paper we establish a MSHT FCLT for the *initial content process* (ICP) of a large-scale queueing system. The ICP specifies the number of customers that were initially in service at time 0 and are still in service later at time t and the elapsed service times since their arrival times before time 0. Assuming that the service times come from a sequence of *independent and identically distributed* (i.i.d.) random variables, independent of the arrival process and system history, the ICP is a Markov process, and thus provides a useful description of the system state at each time. The key assumption is a FCLT for the initial age process, which requires that the number of customers initially in service grows. The technical challenge is treating non-identically distributed remaining service times.

Since MSHT FCLT's for IS models can be fruitfully applied to establish associated MSHT FCLT's for finite-server models [24, 32, 34], our results here have broader implications. In particular, we intend to apply the results here to establish a MSHT FCLT for the $G_t/GI/s_t + GI$ model with time-varying arrival rate and staffing, customer abandonment (the $+GI$) and alternating overloaded (OL) and underloaded (UL) intervals, extending the FCLT for the $G_t/M/s_t + GI$ model in [24]. The present results apply in three ways. First, the theory here applies directly to UL intervals, which can directly be regarded as IS models, starting off with customers in service with elapsed service times, determined from the previous OL interval. Second, the theory here also directly applies to the initial content in service during an OL interval, determined from the previous UL interval (because the dynamics of the ICP is not affected by the finite service capacity). Third, the theory will once again apply to treat the number of waiting customers in an OL interval, because then we can regard the abandonment times as service times; see [24]. There is much more to the total proof, but the present paper provides a key component.

To demonstrate that our assumptions directly cover meaningful cases for IS models, we establish our main MSHT FCLT for the ICP in a more general context. In particular, we consider an IS model in which all arrivals before time 0 have i.i.d. service times with one service-time cdf, while all arrivals after time 0 have i.i.d. service times with another service-time cdf. We refer to the model as $G_t/GI^o, GI^\nu/\infty$ (using the superscript o for old and ν for new). This model represents switching from one kind if service to another in an IS model at time 0.

As usual for FCLT's, we consider a sequence of models indexed by n. For each n, there are infinitely many servers, so that each customer enters service immediately upon arrival. In system n, there is a general arrival process with a time-varying arrival rate function $\lambda_n(t) = n\lambda(t)$ (the G_t), so that the arrival rate is scaled by n, the usual many-server heavy-traffic scaling. We will specify the arrival process only by the requirement that it satisfy an FCLT; see Assumption 1 below.

We assume that the system operated in the past (prior to time 0) as a conventional $G_t/GI/\infty$, model with i.i.d. service times that are independent of the arrival process, distributed according to a *cumulative distribution function* (cdf) G. We assume that new input in the system operates after time 0 according to a $G_t/GI/\infty$ IS model with i.i.d. service times that are independent of the arrival process, with service times distributed according to the cdf G_ν. As in the usual many-server heavy-traffic scaling, the two service-time cdf's G and G_ν are not scaled by n. Our approach is designed especially to treat the case in which these cdf's are different and *not* exponential. An important nontrivial case covered by this $G_t/GI^o, GI^\nu/\infty$ IS model is an IS system starting empty at time $t_0 < 0$. The ICP then describes the state of old content after time 0.

The system performance after time 0 can be characterized by the pair of two-parameter stochastic processes $(X_n^{e,o}(t,y), X_n^{e,\nu}(t,y))$ with $t \geq 0$ and $y \geq 0$. The variable $X^{e,o}(t,y)$ counts the number of customers that were already in service at time 0 and are still in service at time t and have elapsed service times that are less than or equal to y (here $y > t$ since they started service prior to time 0). The variable $X^{e,\nu}(t,y)$ counts the number of customers that arrived after time 0 and are still in service at time t and have elapsed service times that are less than or equal to y (here $0 \le y \le t$ since they started service after to time 0). (The superscripts are chosen to help, with e denoting elapsed, o old and ν new.) Given the assumptions on the service times, the stochastic process $(X_n^{e,o}, X_n^{e,\nu}) \equiv \{(X_n^{e,o}(t,\cdot), X_n^{e,\nu}(t,\cdot)) : t \ge 0\}$ is a Markov process with time domain $[0,\infty)$ and state space $\mathbb{D}^2$, where $\mathbb{D}$ is the usual function space of right-continuous real-valued functions with left limits, endowed with the usual Skorohod topology [41].

Our main result, Theorem 3.2, is an FCLT for $(X_n^{e,o}, X_n^{e,\nu})$ jointly with other processes in the space $\mathbb{D}_{\mathbb{D}^2}$ of $\mathbb{D}^2$-valued functions in $\mathbb{D}$. The use of $\mathbb{D}_{\mathbb{D}^2}$ follows [30, 31, 38]. It is an alternative to measure-valued approaches in [3, 17, 18, 43] and distribution-valued approach in [35]. The alternative approaches are appealing for simplifying arguments and revealing structure; e.g., [35] shows that the the heavy-traffic limit for the $G/G/\infty$ model can be regarded as a tempered-distribution-valued Ornstein-Uhlenbeck diffusion process, generalizing the diffusion process limit for the $M/GI/\infty$ model in [3]. On the other hand, the $\mathbb{D}_{\mathbb{D}}$ framework here evidently admits more continuous functions, and so has more immediate applications via the continuous mapping theorem. Explicit connections between the two approaches for the fluid limits are made in [16].

We contribute here by treating the ICP $X_n^{e,o}$; the limit for $X_n^{e,\nu}$ comes from [30]. As in [30], and in Louchard [26] and Krichagina and Puhalskii [20] before, we work with the empirical process of the service times. As can be seen from §2.2 of [41], §14 of [1] and especially Shorack and Wellner [36], empirical processes and associated statistical tests have been a major focus of FCLT's, ever since [4], so that there are many useful tools for queueing theory.

In particular, to address the technical challenge of non-identically distributed remaining service times, we draw on Chapter 25 of [36], which in turn uses a symmetrization argument from [27], which can be traced back to [39]. Substantial new arguments are required as can be seen from the tightness proof in §4.2.3. Evidently, this is the first use of symmetrization technique to analyze a queueing model with non-identically distributed service times.

We emphasize engineering relevance, e.g., by providing an explicit characterization of the limit process, exposing key structure (see Remark 3.3) and providing explicit formulas for time-varying means, variances and covariances that lead to an effective algorithm for computing relevant performance measures, as confirmed by simulation experiments, which mostly appear in an appendix. on the authors' web pages.

Organization of the paper. In §2, we specify the model operating after time 0 in more detail. We define its key performance functions, specify the MSHT scaling, and our detailed assumptions. At this point, we represent the behavior before time 0 by the assumed behavior of the ICP at time 0 in Assumption 1. In §3 we state our main results and in §4 we prove them. In §5 we show that our results apply to the $G_t/GI^o, GI^\nu/\infty$ model starting some time before time 0, if we assume that the limit for the arrival process has independent increments, because then the ICP at time 0 has the properties assumed in Assumption 1. From this case, we can also see that the results are consistent with the previous results in [30].

Extra materials are given in the appendix. In §A we review a necessary and sufficient condition for tightness in space $\mathbb{D}_{\mathbb{D}}$. In §B we review useful results used in the proofs. In §C we give proofs omitted in the main paper. In §D we report simulation results for a challenging test case having non-Markov arrival process, non-exponential service-time distribution, and general initial conditions. In §E we provide additional simulation results. In §F we provide the first characterization of the steady-state distribution of the new and old content in the stationary $G/GI/\infty$ model.

2. The model. We start by considering the model after time 0; we show that the results can be applied to the $G_t/GI^o, GI^\nu/\infty$ model starting with the initial conditions here at some time before time 0 in §5. Even though we consider time with $t \geq 0$, we are especially interested in those customers who arrived before time 0. Their history will be captured by the initial age process, which coincides with the ICP at $t = 0$.

We are primarily interested in the ICP $X_n^{e,o}(t,y)$, but we also consider the associated process for the new input $X_n^{e,\nu}(t,y)$). In addition to the pair of two-parameter stochastic processes $(X_n^{e,o}(t,y), X_n^{e,\nu}(t,y))$, counting the old and new customers in the system at time t with elapsed service times at most y, we also define the closely related pair of two-parameter stochastic processes $(X_n^{r,o}(t,y), X_n^{r,\nu}(t,y))$, counting the old and new customers in the system at time t with *remaining* service times at least y. Of course, these remaining-time processes are usually not directly observable, but they do usefully represent the future demand. However, they are tightly linked with the other processes. In particular, they are linked via the simple relations $X_n^r(t,y) = X_n(t+y) - X_n^e(t+y,y)$ and $X_n^e(t,y) = X_n(t) - X_n^r(t-y,y)$, where $X_n(t)$ is the total number of customers in system n at time t; i.e., $X_n(t) = X_n^e(t,\infty) = X_n^r(t,0)$.

As indicated in §1, it is important to treat the old and new customers separately. Let $X_n^e(t,y) \equiv X_n^{e,\nu}(t,y) + X_n^{e,o}(t,y)$ and $X_n^r(t,y) \equiv X_n^{r,\nu}(t,y) +$

$X_n^{r,o}(t,y)$. As in [30], for the new arrivals we have

$$(2.1)\qquad X_n^{e,\nu}(t,y) = \sum_{i=N_n((t-y)^+)}^{N_n(t)} \mathbf{1}(A_i^{(n)} + S_i > t), \quad t \geq 0,\ y \geq 0,$$

$$(2.2)\qquad X_n^{r,\nu}(t,y) = \sum_{i=1}^{N_n(t)} \mathbf{1}(A_i^{(n)} + S_i > t + y), \quad t \geq 0,\ y \geq 0$$

where $A_i^{(n)}$ is the arrival time of the i^{th} customer and S_i is the associated service time in system n. The service times S_i has not been scaled by n, hence no superscript.

Now we turn to the processes associated with initial customers already in the system at time 0. Let $\tau_{n,i}$ denote the length of time the i^{th} customer has been in service (age in service) at time 0 in system n. Without loss of generality we assume the ages are ordered $0 \leq \tau_{n,1} \leq \tau_{n,2} \leq \ldots$. Then

$$(2.3)\qquad X_n^{e,o}(t,y) = \sum_{i=1}^{X_n^e(0,(y-t)^+)} \mathbf{1}(\eta_i(\tau_{n,i}) > t), \quad t \geq 0,\ y \geq 0,$$

$$(2.4)\qquad X_n^{r,o}(t,y) = \sum_{i=1}^{X_n(0)} \mathbf{1}(\eta_i(\tau_{n,i}) > t + y), \quad t \geq 0,\ y \geq 0,$$

where $X_n^e(0,(y-t)^+)$ is the total number of customers at time 0 that have been in service for time $(y-t)^+ \equiv \max\{y-t, 0\}$.

The key property we will exploit is the conditional independence property: Conditional on the sequence of service age random variables $\{\tau_{n,i} : i \geq 1\}$, the sequence $\{\eta_i(\tau_{n,i}) : i \geq 1\}$ is a sequence of mutually independent random variables with conditional tail probabilities

$$(2.5)\qquad P(\eta_i(x) > t | \tau_{n,i} = x) \equiv H_x^c(t) \equiv 1 - H_x(t) \equiv \frac{G^c(t+x)}{G^c(x)},$$

for $x \geq 0$, $t \geq 0$, where $G^c(x) \equiv 1 - G(x)$ is the *complementary cdf* (ccdf) for the service distribution of old customers. The primary difficulty in the proof stems from the fact that, conditional on the sequence of service age random variables $\{\tau_{n,i} : i \geq 1\}$, the random variables $\eta_i(\tau_{n,i})$ are *not identically distributed*.

Given the processes (2.1)-(2.4) and the equalities $X_n(t) = X_n^e(t,\infty) = X_n^r(t,0)$, we can define the departure process associated with initial and new customers from the n^{th} queue. Let $D_n^o(t)$ ($D_n^\nu(t)$) be the total number of initial (new) customers who have departed by time t. Then necessarily $D_n^o(t) = X_n(0) - X_n^o(t)$ and $D_n^\nu(t) = N_n(t) - X_n^\nu(t)$. Hence $D_n(t) \equiv D_n^o(t) + D_n^\nu(t) = X_n(0) + N_n(t) - X_n(t)$ represents the total number of departures by time t.

Associated scaled processes. Let the associated LLN-scaled processes be

$$\begin{aligned} \bar{N}_n(t) &\equiv N_n(t)/n, \quad \bar{X}_n^e(t,y) \equiv X_n^e(t,y)/n, \\ \bar{D}_n(t) &\equiv D_n(t)/n, \quad \bar{X}_n^r(t,y) \equiv X_n^r(t,y)/n. \end{aligned} \tag{2.6}$$

Let the associated CLT-scaled processes be

$$\begin{aligned} \hat{N}_n(t) &\equiv \frac{N_n(t) - n\Lambda(t)}{\sqrt{n}}, \quad \hat{X}_n^e(t,y) \equiv \frac{X_n^e(t,y) - nX^e(t,y)}{\sqrt{n}}, \\ \hat{D}_n(t) &\equiv \frac{D_n(t) - nD(t)}{\sqrt{n}}, \quad \hat{X}_n^r(t,y) \equiv \frac{X_n^r(t,y) - nX^r(t,y)}{\sqrt{n}}, \end{aligned} \tag{2.7}$$

where the centering terms $\Lambda(t)$, $X^e(t,y)$, $X^r(t,y)$, $D(t)$ are deterministic functions (fluid limits) to be specified below in Assumption 1 and Theorem 3.1.

The spaces $\mathbb{D}$ and $\mathbb{D}_{\mathbb{D}}$. The limits are established in the function space $\mathbb{D} \equiv \mathbb{D}([0,\infty),\mathbb{R})$ of right continuous functions with left limits equipped with the Skorohod J_1 topology and the associated metric d_{J_1} [6, 15, 37, 41]. Products of that space are equipped with the product topology. Since all limits will almost surely have continuous sample paths, convergence in J_1 topology is equivalent to uniform convergence over compact sets (time intervals). For the two-parameter processes, the processes are random elements of the space $\mathbb{D}_{\mathbb{D}} \equiv \mathbb{D}([0,\infty),\mathbb{D}([0,\infty),\mathbb{R}))$ of $\mathbb{D}$-valued functions. Since the space $(\mathbb{D}, J_1)$ is a complete separable metric space, this space of $\mathbb{D}$-valued functions falls within Skorohod's [37] original framework; see [30, 38] for more details. We prove convergence in these spaces by using the compactness approach, i.e., by proving convergence of the *finite dimensional distribution* (fidis) and tightness of the processes; see [1, 6, 15, 41] for tightness criteria in $\mathbb{D}$ and Theorem 6.2 of [30] for tightness criteria in $\mathbb{D}_{\mathbb{D}}$. We review the tightness criteria in $\mathbb{D}_{\mathbb{D}}$ in §A.1.

Assumptions. Our key assumption is a joint FCLT for the arrival process of new customers after time 0 and for the initial ages. We discuss the appropriateness of this assumption in Remarks 2.3 and 2.4 below and in §5.

ASSUMPTION 1 (Joint FCLT for the arrival process and initial ages). *The CLT-scaled ICP and external arrival processes defined in* (2.7) *jointly* *satisfy the FCLT*

$$\left(\hat{X}^e_n(0,\cdot), \hat{N}_n\right) \Rightarrow \left(\hat{X}^e(0,\cdot), \hat{N}\right) \quad \textit{in} \quad \mathbb{D}^2 \quad \textit{as} \quad n \to \infty, \tag{2.8}$$

where $\hat{X}^e_n(0,\cdot)$ *and* $\hat{N}$ *are two independent zero-mean continuous Gaussian processes. We assume that the deterministic centering terms in* (2.7)*, which come from the associated functional weak law of large numbers* (*FWLLN*) *stated below in* (2.10)*, can be represented as*

$$X^e(0,x) = \int_0^x a(u)\,du, \quad x \geq 0, \quad \textit{and} \quad \Lambda(t) = \int_0^t \lambda(u)\,du, \quad t \geq 0, \tag{2.9}$$

where the fluid initial age density $a(x)$ *and arrival rate function* $\lambda(t)$ *in* (2.9) *are nonnegative real-valued functions that are integrable over all bounded intervals.*

Remark 2.1 (FWLLN for the arrival process and initial content in service). As an immediate consequence of Assumption 1, we have a FWLLN for $\bar{N}_n$ and $\bar{X}^e_n(0,\cdot)$, i.e., as $n \to \infty$,

$$\left(\bar{X}^e_n(0,\cdot), \bar{N}_n, \bar{X}_n(0)\right) \Rightarrow (X^e(0,\cdot), \Lambda, X(0)) \quad \text{in} \quad \mathbb{D}^2 \times \mathbb{R}. \tag{2.10}$$

Remark 2.2 (The zero-mean Gaussian assumption). The zero-mean Gaussian requirement of Assumption 1 is not required for the convergence, but it is required for drawing the useful conclusion that the limit process also has this structure, as in (3.4) below. Nevertheless, the assumption is natural. Extensions are possible, as illustrated by §10 of [24].

Remark 2.3 (joint convergence and independence of the limits). If the arrival process is a nonhomogeneous Poisson process, so that the IS model becomes $M_t/GI^o, GI^\nu/\infty$, then the new input after time 0 is independent of the initial content, so that the independence of the two limit processes follows directly from the two separate limits in Assumption 1. But, more generally, the number of customers in service at time 0 and the ages of the service times of those customers typically will *not* be independent of the arrivals after time 0. Thus, Assumption 1 may not be easy to verify. Nevertheless, Assumption 1 is very reasonable. It is what we expect to be true in great generality. For example, consider a $G_t/GI/\infty$ system starting empty in the finite past. Even though the arrival process may not have independent increments, from [30] we know that it is common for the limit of the arrival process to be a time-transformed *Brownian motion* (BM), which has independent increments. In particular, that occurs if we assume that the arrival process is a deterministic time transformation of any arrival process that satisfies an FCLT with a BM limit. For such limits, it is natural to start with a stationary process, such as an equilibrium renewal process, but it suffices to have the FCLT with a BM limit, as discussed in §7 of [28]. With either an ordinary or equilibrium renewal process, the limit process will be $\hat{N}(t) = c_\lambda \mathcal{B}_a(\Lambda(t))$, where $\mathcal{B}_a$ is a standard BM, $\Lambda(t)$ is the deterministic time transformation, corresponding to the limiting cumulative arrival rate function and c_λ^2 is the *squared coefficient of variation* (SCV, variance divided by the square of the mean) of an interarrival time in the ordinary renewal process. For all these representations, the arrival FCLT and the independence of the limit is satisfied, as assumed in Assumption 1. In addition to [28], see [8, 11, 21, 25] for uses of this representation of nonstationary non-Poisson arrival processes.

REMARK 2.4 (Performance forecasting using limits in Assumption 1). For engineering purposes, the limits in Assumption 1 can be understood as estimators (approximations) for future demand posed by new input and initial content. The goal here is to develop performance forecasting formulas as functions of the limits in Assumption 1. It will be clear from the formulas and examples that the general initial conditions (represented by the initial fluid age function $X^e(0,\cdot)$ and the associated stochastic limit process $\hat{X}^e(0,\cdot)$) can be a significant part of the performance functions. □

We also impose additional regularity assumptions, which evidently are not too restrictive for engineering applications. We first impose conditions on the two service-time cdf's. Even though not restrictive, both assumptions are used critically in the analysis; see Remark 3.2 and Lemma 4.4.

ASSUMPTION 2 (Regularity conditions for service-time cdf's). *The two service-time cdf's G and G_ν are assumed to be continuous. In addition, the cdf G has a probability density function (pdf) g satisfying $0 < g(x) \le g^\uparrow \equiv \sup_{x \ge 0} g(x) < \infty$ for all $x \ge 0$.*

We also impose a regularity condition on the initial content. It is used in the proof of tightness in $\mathbb{D}_\mathbb{D}$ in §4.2.3.

ASSUMPTION 3 (Regularity conditions for the initial content). *We assume that there exists $y^\uparrow > 0$ such that $X_n(0) - X_n(0, y^\uparrow) = 0$ for all $n \ge 1$ w.p.1..*

3. Main results. In this section, we present the new FWLLN and FCLT for the $G_t/GI^o, GI^\nu/\infty$ model. They extend the corresponding results for the $G_t/GI/\infty$ model in §3 and §5 of [30] by treating more general initial conditions. In particular, the results for the new arrivals come from [30], but unlike §5 of [30], Assumption 1 here makes the remaining service times at time 0 be conditionally independent, given the ages, but not identically distributed random variables. We state the FWLLN first, but give no separate proof, because it is a consequence of the FCLT.

THEOREM 3.1 (FWLLN). *Consider the sequence of $G_t/GI^o, GI^\nu/\infty$ queues satisfying all assumptions in §2. As $n \to \infty$,*

$$\big(\bar{N}_n, \bar{X}^e_n(0,\cdot), \bar{X}^r_n(0,\cdot), \bar{X}^e_n, \bar{X}^r_n, \bar{X}_n, \bar{D}_n\big) \Rightarrow (\Lambda, X^e(0,\cdot), X^r(0,\cdot), X^e, X^r, X, D) \tag{3.1}$$

in $\mathbb{D}^3 \times \mathbb{D}^2_{\mathbb{D}} \times \mathbb{D}^2$, where the limit is continuous and deterministic with $X(t) = X^e(t,\infty) = X^r(t,0)$, and

$$X^e(t,y) \equiv X^{e,o}(t,y) + X^{e,\nu}(t,y), \quad X^r(t,y) \equiv X^{r,o}(t,y) + X^{r,\nu}(t,y),$$

$$X^{e,o}(t,y) = \int_0^{(y-t)^+} a(x) H^c_x(t)dx, \quad X^{e,\nu}(t,y) \equiv \int_{(t-y)^+}^t G^c_\nu(t-s)\lambda(s)\,ds,$$

$$X^{r,o}(t,y) = \int_0^\infty a(x) H^c_x(t+y)dx, \quad X^{r,\nu}(t,y) \equiv \int_0^t G^c_\nu(t+y-s)\lambda(s)\,ds,$$

$$D(t) = \Lambda(t) - X(t) = \int_0^\infty a(x) H_x(t)\,dx + \int_0^t G_\nu(t-s)\lambda(s)\,ds \tag{3.2}$$

and $a(x)$ being the initial fluid limit age density and $\lambda(s)$ being the arrival rate function specified in Assumption 1.

For real numbers a and b, let $a \vee b \equiv \max\{a,b\}$ and $a \wedge b \equiv \min\{a,b\}$.

THEOREM 3.2 (FCLT). *Consider the sequence of $G_t/GI^o, GI^\nu/\infty$ IS models satisfying all assumptions in §2. As $n \to \infty$,*

$$\begin{aligned}\Big(\hat{N}_n, \hat{X}^e_n(0,\cdot), \hat{X}^r_n(0,\cdot), \hat{X}^e_n, \hat{X}^r_n, \hat{X}_n, \hat{D}_n\Big) \\ \Rightarrow \Big(\hat{N}, \hat{X}^e(0,\cdot), \hat{X}^r(0,\cdot), \hat{X}^e, \hat{X}^r, \hat{X}, \hat{D}\Big)\end{aligned} \tag{3.3}$$

in $\mathbb{D}^3 \times \mathbb{D}^2_{\mathbb{D}} \times \mathbb{D}^2$, where the stochastic limit process for the two-parameter ICP, the scaled number of customers in service at t with age at most y, is

$$\hat{X}^e(t,y) = \hat{X}^{e,\nu}_1(t,y) + \hat{X}^{e,\nu}_2(t,y) + \hat{X}^{e,o}_1(t,y) + \hat{X}^{e,o}_2(t,y), \tag{3.4}$$

where $\hat{X}_1^{e,\nu}$, $\hat{X}_2^{e,\nu}$, $\hat{X}_1^{e,o}$ and $\hat{X}_2^{e,o}$ are independent zero-mean Gaussian processes with continuous sample paths,

$$\hat{X}_1^{e,\nu}(t,y) \equiv \int_{(t-y)^+}^{t} G_\nu^c(t-s)\,d\hat{N}(s), \tag{3.5}$$

$$\hat{X}_2^{e,\nu}(t,y) \equiv \int_{(t-y)^+}^{t} \int_0^\infty \mathbf{1}(x > t-s)\,d\hat{K}_\nu(\Lambda(s),x), \tag{3.6}$$

where $\hat{N}$ is the limit process in the assumed FCLT for the arrival process specified in Assumption 1, and $\hat{K}_\nu(t,x) \equiv \hat{U}(t,G(x))$, with $\hat{U}$ being a standard Kiefer process, capturing the variability of the new service times, and independent of $\hat{N}$; $\hat{X}_1^{e,o}$ is a zero-mean Gaussian process with the covariance function

$$\begin{aligned} C_1^{e,o}((t_1,y_1),(t_2,y_2)) &\equiv \mathrm{Cov}\left(\hat{X}_1^{e,o}(t_1,y_1), \hat{X}_1^{e,o}(t_2,y_2)\right) \\ &= \int_0^{(y_1-t_1)^+\wedge(y_2-t_2)^+} H_u(t_1\wedge t_2) H_u^c(t_1\vee t_2)\,dX^e(0,u), \end{aligned} \tag{3.7}$$

and $\hat{X}_2^{e,o}$ has the representation

$$\begin{aligned} \hat{X}_2^{e,o}(t,y) &\equiv \int_0^{(y-t)^+} H_x^c(t)\,d\hat{X}^e(0,x) \\ &\equiv H_{(y-t)^+}^c(t)\hat{X}^e(0,(y-t)^+) - \int_0^{(y-t)^+} \hat{X}^e(0,u-)dH_u^c(t), \end{aligned} \tag{3.8}$$

where $(X^e(0,\cdot),\hat{X}^e(0,\cdot))$ is the limit of the initial age process in Assumption 1 and (2.10). The joint limit (3.3) follows from the displayed limit. The other limit processes $\hat{X}$, $\hat{D}$ and $\hat{X}^r$ are specified in the corollaries below.

Remark 3.1 (Correction in [30]). The limits for the new input follow from [30], so the formulas in (3.5) and (3.6) should be consistent with [30]. However, here we make a correction, noting that the upper limit of the inner integrals in (2.10), (2.15) and for $X_2^{c,e}(t,y)$ in (3.16) of [30] all should be ∞ instead of t. Similarly the upper limit of the second integral in the expression for $\sigma_{q,e}^2(t,y)$ in Theorem 4.2 of [30] also should be ∞ instead of t. After this correction, the formulas in (3.6) and elsewhere are consistent with [30]. □

We next characterize all the other limit processes using the limit in (3.4). Let $\stackrel{\mathrm{d}}{=}_t$ denote equal in distribution for each t. Let $\mathcal{B}_s(\cdot)$ be an independent BM (associated with service times of new customers).

COROLLARY 3.1 (Limits for the one-parameter queue length process). *Under the assumptions of Theorem 3.2, the limit for the total number in service at t is*

$$\hat{X}(t) \equiv \hat{X}^e(t,\infty) \equiv \hat{X}_1^\nu(t) + \hat{X}_2^\nu(t) + \hat{X}_1^o(t) + \hat{X}_2^o(t), \tag{3.9}$$

where $\hat{X}_1^\nu$, $\hat{X}_2^\nu$, $\hat{X}_1^o$ *and* $\hat{X}_2^o$ *are independent zero-mean Gaussian processes with continuous sample paths and*

$$\hat{X}_1^\nu(t) \equiv \hat{X}_1^{e,\nu}(t,\infty) \equiv \int_0^t G_\nu^c(t-s)\, d\hat{N}(s),, \tag{3.10}$$

$$\begin{aligned}\hat{X}_2^\nu(t) \equiv \hat{X}_2^{e,\nu}(t,\infty) &\equiv \int_0^t \int_0^\infty \mathbf{1}(x > t-s)\, d\hat{K}_\nu(\Lambda(s),x) \\ &\stackrel{\mathrm{d}}{=}_t - \int_0^t \sqrt{G_\nu(t-s)G_\nu^c(t-s)} d\mathcal{B}_s(\Lambda(s)),\end{aligned} \tag{3.11}$$

$\hat{X}_1^o(t) \equiv \hat{X}_1^{e,o}(t,\infty)$ *is a zero-mean Gaussian process with the covariance function*

$$\begin{aligned} C_1^o(t,t') &\equiv \mathrm{Cov}\left(\hat{X}_1^o(t), \hat{X}_1^o(t')\right) \\ &= C_1^{e,o}((t_1,\infty),(t_2,\infty)) = \int_0^\infty H_u(t\wedge t')H_u^c(t\vee t')\, dX^e(0,u),\end{aligned} \tag{3.12}$$

and

$$\hat{X}_2^o(t) \equiv \hat{X}_2^{e,o}(t,\infty) \equiv \int_0^\infty H_x^c(t)\, d\hat{X}^e(0,x). \tag{3.13}$$

COROLLARY 3.2 (Limits for the one-parameter departure process). *Under the assumptions of Theorem 3.2, The limit for the number of departures by t is*

$$\hat{D}(t) = \hat{D}_1^\nu(t) + \hat{D}_2^\nu(t) + \hat{D}_1^o(t) + \hat{D}_2^o(t), \tag{3.14}$$

where $\hat{D}_1^\nu$, $\hat{D}_2^\nu$, $\hat{D}_1^o$ *and* $\hat{D}_2^o$ *are independent zero-mean Gaussian processes, with*

$$\hat{D}_1^\nu(t) \equiv \int_0^t G_\nu(t-s)\, d\hat{N}(s), \tag{3.15}$$

$$\begin{aligned}\hat{D}_2^\nu(t) &\equiv \int_0^t \int_0^\infty \mathbf{1}(x \le t-s)\, d\hat{K}_\nu(\Lambda(s),x) \\ &\stackrel{\mathrm{d}}{=}_t \int_0^t \sqrt{G_\nu(t-s)G_\nu^c(t-s)} d\mathcal{B}_s(\Lambda(s)),\end{aligned} \tag{3.16}$$

$\hat{D}_1^o(t) = -\hat{X}_1^o(t)$ *being a zero-mean Gaussian process with covariance function* $\mathrm{Cov}\left(\hat{D}_1^o(t), \hat{D}_1^o(t')\right) = C_1^o(t,t')$, *and*

$$\hat{D}_2^o(t) \equiv \hat{X}(0) - \hat{X}_2^o(t) = \int_0^\infty H_x(t)\, d\hat{X}^e(0,x). \tag{3.17}$$

COROLLARY 3.3 (Limits for the remaining-service-time process). *Under the assumptions of Theorem 3.2, the limit* $\hat{X}^r(0,x) = \hat{X}^o(x) = \hat{X}_1^{e,o}(x,\infty) + \hat{X}_2^{e,o}(x,\infty)$ *for all* $x \geq 0$ *and*

$$\hat{X}^r(t,x) = \hat{X}_1^{r,\nu}(t,x) + \hat{X}_2^{r,\nu}(t,x) + \hat{X}_1^{r,o}(t,x) + \hat{X}_2^{r,o}(t,x), \tag{3.18}$$

with $\hat{X}_1^{r,\nu}$, $\hat{X}_2^{r,\nu}$, $\hat{X}_1^{r,o}$ *and* $\hat{X}_2^{r,o}$ *being independent zero-mean Gaussian processes and*

$$\hat{X}_1^{r,\nu}(t,x) \equiv \int_0^t G_\nu^c(t+x-s)\, d\hat{N}(s), \tag{3.19}$$

$$\begin{aligned} \hat{X}_2^{r,\nu}(t,x) &\equiv \int_0^t \int_0^\infty \mathbf{1}(u+s > t+x)\, d\hat{K}_\nu(\Lambda(s),u), \\ \hat{X}_1^{r,o}(t,x) &\equiv \hat{X}_1^o(t+x) = \hat{X}_1^{e,o}(t+x,\infty) \quad \textit{and} \\ \hat{X}_2^{r,o}(t,x) &\equiv \hat{X}_2^o(t+x) = \hat{X}_2^{e,o}(t+x,\infty). \end{aligned} \tag{3.20}$$

REMARK 3.2 (The stochastic integrals). The integrals in Theorem 3.2 should be interpreted just as in [30], as explained in Remark 3.2 there. In particular, the deterministic integrals in (3.7) and (3.12) are all Stieltjes integrals, while the integrals in (3.6), (3.11), (3.16) and (3.20) are two-parameter stochastic integrals, just as in [23, 24, 30]. As in Theorem 3.2 and Remark 3.3 of [30], the continuity assumption on the cdf G_ν in Assumption 2 is used to get the representation in terms of the Kiefer process.

Of special note are the stochastic integrals with respect to $\hat{N}$ in (3.5), (3.10), (3.15) and (3.19), and with respect to $\hat{X}^e(0,\cdot)$ in (3.8), (3.13) and (3.17). As explained in Remark 3.2 of [30], these all should be interpreted as the form after the representation of integration by parts, as given on p. 336 of [2]. That is justified because the pre-limit processes of the integrator process have sample paths of bounded variation. For example, the alternative representation for (3.8) is given there; see §4.3. Finally, the stochastic integral with respect to $\hat{K}_\nu$ should be understood in the mean-square sense, as in §6.3 of [20].

REMARK 3.3 (Four independent stochastic effects). The expression for the limit process $\hat{X}^e$ in (3.4) as the sum of the four independent processes

$\hat{X}_1^{e,\nu}$, $\hat{X}_2^{e,\nu}$, $\hat{X}_1^{e,o}$ and $\hat{X}_2^{e,o}$ shows that the four sources of variability in the model contribute to the total variability independently. The process $\hat{X}_1^{e,\nu}$ captures the variability in the arrival process after time 0; the process $\hat{X}_2^{e,\nu}$ captures the variability in the service times after time 0; the process $\hat{X}_1^{e,o}$ captures the variability in the remaining service times at time 0 given that the initial age process is around $nX^e(0,\cdot)$; and the process $\hat{X}_2^{e,o}$ captures the variability of the ages of initial customers at time 0. It is easy to see that this separation of variability effects is not baked in, because this separation does not apply to the pre-limit processes. In the limits for these terms, the impact from other model components become deterministic. We remark that this kind of separation of variability has been observed in the past. For instance, in [24], the FCLT limit of the waiting time process solves a stochastic differential equation driven by three independent BMs that are associated with the arrival process, abandonment times and service process. □

If we make an additional assumption for $\hat{N}$, we can exhibit covariance and variance formulas.

COROLLARY 3.4 (Variance and covariance formulas for the ICP). *If $\hat{N}(t) = c_a \mathcal{B}_a(\Lambda(t))$ and $\Sigma_2^{e,o}(t) \equiv Var(\hat{X}^e(0,t))$, then the covariances of $\hat{X}^e$ the covariances are*

$$\begin{aligned} C^e((t_1,y_1),(t_2,y_2)) &\equiv \mathrm{Cov}(\hat{X}^e(t_1,y_1),\hat{X}^e(t_2,y_2)) \\ &= C^{e,\nu}((t_1,y_1),(t_2,y_2)) + C^{e,o}((t_1,y_1),(t_2,y_2)), \end{aligned}$$

where

$$\begin{aligned} C^{e,\nu}((t_1,y_1),(t_2,y_2)) &\equiv \int_{(t_1-y_1)^+\vee(t_2-y_2)^+}^{t_1\wedge t_2} \big[(c_a^2-1)G_\nu^c(t_1-s)G_\nu^c(t_2-s) \\ &\quad + G_\nu^c((t_1\vee t_2)-s)\big]\lambda(s)ds, \\ C^{e,o}((t_1,y_1),(t_2,y_2)) &\equiv \int_0^{(y_1-t_1)^+\wedge(y_2-t_2)^+} H_u(t_1\wedge t_2)H_u^c(t_1\vee t_2)\,dX^e(0,u) \\ &\quad + \int_0^{(y_1-t_1)^+\wedge(y_2-t_2)^+} H_u^c(t_1)H_u^c(t_2)\,d\Sigma_2^{e,o}(u). \end{aligned}$$

so that the variances are

$$\begin{aligned} \sigma_e^2(t,y) &\equiv \mathrm{Var}(\hat{X}^e(t,y)) = \sigma_{e,\nu}^2(t,y) + \sigma_{e,o}^2(t,y), \\ &\textit{where}\quad \sigma_{e,\nu}^2(t,y) = \sigma_\nu^2(((t-y)^+,t), \end{aligned}$$

$$\sigma^2_\nu(u,v) \equiv \int_u^v \left[(c_a^2-1)G^c_\nu(v-s)^2 + G^c_\nu(v-s)\right]\lambda(s)ds$$
$$\text{and} \quad \sigma^2_{e,o}(t,y) = \int_0^{(y-t)^+} H_u(t)H^c_u(t)\,dX^e_0(u) + \int_0^{(y-t)^+} H^c_x(t)^2\,d\Sigma^{e,o}_2(x).$$

COROLLARY 3.5 (Variance for $\hat{X}(t)$ and $\hat{D}(t)$). *Under the assumptions of Corollary 3.4, the variances of the one-parameter processes $\hat{X}(t)$ and $\hat{D}(t)$ are*

$$\sigma^2_{\hat{X}}(t) \equiv \mathrm{Var}(\hat{X}(t)) = \sigma^2_{\hat{X},\nu}(t) + \sigma^2_{\hat{X},o}(t), \tag{3.21}$$

where

$$\sigma^2_{\hat{X},\nu}(t) \equiv \sigma^2_{e,\nu}(t,\infty) = \sigma^2_\nu(0,t) = \int_0^t \left[(c_a^2-1)G^c_\nu(t-s)^2 + G^c_\nu(t-s)\right]\lambda(s)ds \tag{3.22}$$

and

$$\sigma^2_{\hat{X},o}(t) = \sigma^2_{e,o}(t,\infty) = \int_0^\infty H_u(t)H^c_u(t)\,dX^e_0(u) + \int_0^\infty H^c_x(t)^2\,d\Sigma^{e,o}_2(x), \tag{3.23}$$

$$\sigma^2_{\hat{D}}(t) \equiv \mathrm{Var}(\hat{D}(t)) = \sigma^2_{\hat{D},\nu}(t) + \sigma^2_{\hat{D},o}(t), \tag{3.24}$$

where

$$\sigma^2_{\hat{D},\nu}(t) = \int_0^t \left[(c_a^2-1)G^2_\nu(t-s) + G_\nu(t-s)\right]\lambda(s)ds$$
$$\text{and} \quad \sigma^2_{\hat{D},o}(t) = \int_0^\infty H_u(t)H^c_u(t)\,dX^e_0(u) + \int_0^\infty H_x(t)^2\,d\Sigma^{e,o}_2(x),$$

REMARK 3.4 (Additivity of the variance formulas). The first term of the variance formula of $\hat{X}(t)$ ($\hat{D}(t)$) $\sigma^2_{\hat{X},\nu}(t)$ ($\sigma^2_{\hat{D},\nu}(t)$) provides the variance when the system is initially empty (which coincides with the variance formula in [30]). The second term $\sigma^2_{\hat{X},o}(t)$ ($\sigma^2_{\hat{D},o}(t)$) represents the variance of the content that has been in the system since time 0.

4. Proof of Theorem 3.2. We start with the FCLT for all processes related to new arrivals from [30], obtaining

$$(\hat{N}_n, \bar{N}_n, \hat{K}_n, \bar{K}_n, \hat{R}_n, \hat{X}^{e,\nu}_n, \hat{X}^{r,\nu}_n, \hat{D}^\nu_n) \Rightarrow (\hat{N}, \bar{N}, \hat{K}, \bar{K}, \hat{R}, \hat{X}^{e,\nu}, \hat{X}^{r,\nu}, \hat{D}^\nu), \tag{4.1}$$

in $\mathbb{D}^3 \times \mathbb{D}^5_{\mathbb{D}}$. By Assumption 1, it remains to show the convergence

$$(\hat{X}_n(0,\cdot), \bar{X}_n(0,\cdot), \hat{X}^{e,o}_n, \hat{X}^{r,o}_n, \hat{D}^o_n) \Rightarrow (\hat{X}(0,\cdot), \bar{X}(0,\cdot), \hat{X}^{e,o}, \hat{X}^{r,o}, \hat{D}^o). \tag{4.2}$$

in $\mathbb{D}^3 \times \mathbb{D}^2_{\mathbb{D}}$. We will then have the joint convergence of (4.1) and (4.2) in $\mathbb{D}^7 \times \mathbb{D}^8_{\mathbb{D}}$. (The joint convergence of $\hat{X}_n$ and $\hat{D}_n$ follows from continuous mapping theorem for addition at continuous limits.) In §4.1 we show that the main two-parameter process $\hat{X}^{e,o}_n$ can be decomposed into two other two-parameter processes $\hat{X}^{e,o}_{n,1}(t,y)$ and $\hat{X}^{e,o}_{n,2}(t,y)$, that can be treated separately by conditioning on the ages at time 0. We establish convergence for those two processes in §§4.2–4.3; In §4.4 we prove the convergence of other processes.

4.1. *Decomposition of $\hat{X}^{e,o}_n$.* To prove (4.2), we use a convenient representation of $\hat{X}^{e,o}_n(t,y)$. Let $\widehat{\eta}_{n,i}(t) \equiv \mathbf{1}\{\eta_i(\tau_{n,i}) > t\} - H^c_{\tau_{n,i}}(t)$. From (2.3), (2.7) and (3.2), we can write

$$\begin{aligned}
\hat{X}^{e,o}_n(t,y) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{X^e_n(0,(y-t)^+)} \mathbf{1}\{\eta_i(\tau_{n,i}) > t\} - \int_0^{(y-t)^+} a(x)H^c_x(t)dx\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{X^e_n(0,(y-t)^+)} \left(\mathbf{1}\{\eta_i(\tau_{n,i}) > t\} - H^c_{\tau_{ni}}(t)\right) \\
&\quad + \sqrt{n}\left(\int_0^{(y-t)^+} H^c_x(t)d\bar{X}^e_n(0,x) - \int_0^{(y-t)^+} a(x)H^c_x(t)dx\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{X^e_n(0,(y-t)^+)} \widehat{\eta}_{n,i}(t) + \int_0^{(y-t)^+} H^c_x(t)d\hat{X}^e_n(0,x) \\
&\equiv \hat{X}^{e,o}_{n,1}(t,y) + \hat{X}^{e,o}_{n,2}(t,y), \qquad (4.3)
\end{aligned}$$

where the second equality holds by adding and subtracting $H_{\tau_{n,i}}(t)$ in the sum.

To prove the convergence of (4.3), we will show the joint convergence of the two terms on the right-hand side of (4.3) and apply the continuous mapping theorem with addition. We know that joint convergence of two random elements is equivalent to the individual convergence of both terms if they are independent. Even though $\hat{X}^{e,o}_{n,1}$ and $\hat{X}^{e,o}_{n,2}$ in (4.3) are not independent, because they both involve the age sequence $\{\tau_{n,j} : j \geq 1\}$ or equivalently the counting process $X^e_n(0,\cdot)$, they are conditionally independent given $\bar{X}^e_n(0,\cdot)$. Hence, in order to treat the two terms separately we condition upon the age sequence and then uncondition. In doing so, we apply the assumed convergence in Assumption 1 together with the following lemma, which expresses the argument used in the proof of Theorem 7.6 of [29], which follows §5 of [42]. In particular, a variant of the following lemma is uses in §7.3 of [29] to extend FCLT's for $M/M/n/m_n$ queues to $G/M/n/m_n$ queues, allowing a general arrival process that satisfies a FCLT. The spaces are different here, but the argument is the same.

LEMMA 4.1. *Let $\{Y_n : n \geq 1\}$ and Y be processes with sample paths in $\mathbb{D}_{\mathbb{D}}$, and let $\{Z_n : n \geq 1\}$ and Z be processes with sample paths in $\mathbb{D}$. Let $Y_n^{Z_n}$ (Y^Z) denote Y_n (Y) conditioned on Z_n (Z). If $Z_n \Rightarrow Z$ in $\mathbb{D}$ and*

$$Y_n^{Z_n} \Rightarrow Y^Z \quad \text{in} \quad \mathbb{D}_{\mathbb{D}} \quad \text{whenever} \quad Z_n \to Z \quad \text{in} \quad \mathbb{D} \quad \text{as} \quad n \to \infty \quad \text{w.p.1}, \tag{4.4}$$

then $Y_n \Rightarrow Y$ in $\mathbb{D}_{\mathbb{D}}$ as $n \to \infty$.

We apply Lemma 4.1 with the initial age process $\hat{X}_n^e(0, \cdot)$ playing the role of Z_n. The required convergence in distribution holds by Assumption 1. We will then condition on the ages and assume that

$$\hat{X}_n^e(0, \cdot) \to \hat{X}^e(0, \cdot) \quad \text{in} \quad \mathbb{D} \quad \text{w.p.1.} \tag{4.5}$$

In (4.5) we use the Skorohod representation theorem to replace convergence in distribution by convergence w.p.1. It remains to establish the limit (4.4), assuming (4.5).

Given that we condition with respect to the ages and then uncondition, in order to establish the joint convergence $\left(\hat{X}_{n,1}^{e,o}, \hat{X}_{n,2}^{e,o}, \hat{X}_n^e(0, \cdot), \bar{X}_n^e(0, \cdot)\right) \Rightarrow \left(\hat{X}_1^{e,o}, \hat{X}_2^{e,o}, \hat{X}_0^e, X_0^e\right)$ in $\mathbb{D}_{\mathbb{D}}^2 \times \mathbb{D}^2$, it suffices to prove $\left(\hat{X}_{n,1}^{e,o}, \bar{X}_n^e(0, \cdot)\right) \Rightarrow \left(\hat{X}_1^{e,o}, X_0^e\right)$ in $\mathbb{D}_{\mathbb{D}} \times \mathbb{D}$ and $\left(\hat{X}_{n,2}^{e,o}, \hat{X}_n^e(0, \cdot), \bar{X}_n^e(0, \cdot)\right) \Rightarrow \left(\hat{X}_2^{e,o}, \hat{X}_0^e, X_0^e\right)$ in $\mathbb{D}_{\mathbb{D}} \times \mathbb{D}^2$; i.e., it suffices to treat the two terms separately. Aside from the conditioning, we would be using Theorems 11.4.4 and 11.4.5 in [41], which justify joint convergence. We next separately prove the convergence of two terms in (4.3).

4.2. *Convergence of the First Term in* (4.3). In addition to the conditioning discussed above, we use the compactness approach to prove (4.4) in order to establish convergence of the first term in (4.3); i.e., we prove convergence of the fidis in $\mathbb{D}_{\mathbb{D}}$ in two steps and then we prove tightness in the third step. In Step 1 (§4.2.1), we establish convergence of the four-parameter covariance functions of $\hat{X}_{n,1}^{e,o}$, referred to as $K_n(t, y, t', y')$, to those of $\hat{X}_1^{e,o}$, defined as $K(t, y, t', y')$ in (3.7) in Theorem 3.2. In Step 2 (§4.2.2), using the convergence of the covariance functions, we establish the convergence of the fidis of $\hat{X}^{e,o}_{n,1}$ in $\mathbb{D}_{\mathbb{D}}$, which is equivalent to the joint convergence of $\left(\hat{X}^{e,o}_{n,1}(t_1,\cdot),\ldots,\hat{X}^{e,o}_{n,1}(t_k,\cdot)\right)$ in $\mathbb{D}^k$, for all $k \geq 1$ and $0 < t_1 < \cdots < t_k$. We do this in two sub-steps: First, we show the convergence of the fidis of the vector $\left(\hat{X}^{e,o}_{n,1}(t_1,\cdot),\ldots,\hat{X}^{e,o}_{n,1}(t_k,\cdot)\right)$ in the second argument, namely, the joint convergence of the bigger vector $\left(\hat{X}^{e,o}_{n,1}(t_i,y_j), 1 \leq i \leq k, 1 \leq j \leq m\right)$ in $\mathbb{R}^{k\times m}$, for all $m \geq 1$ and $0 < y_1 < \cdots < y_m$. Second, we establish the tightness of $\left(\hat{X}^{e,o}_{n,1}(t_1,\cdot),\ldots,\hat{X}^{e,o}_{n,1}(t_k,\cdot)\right)$ in $\mathbb{D}^k$. In Step 3 (§4.2.3) we prove that $\hat{X}^{e,o}_{n,1}$ is tight in $\mathbb{D}_{\mathbb{D}}$.

4.2.1. *Step 1: Convergence of covariance functions.* As indicated above, we start by conditioning on the ages. Let $\mathbb{E}^{\tau}$ denote the conditional expectation operator, conditional on the ages or upon the process $\bar{X}^e_n(0,\cdot)$. Upon conditioning, the first term in (4.3) is a non-random sum of the independent mean-zero random variables $\widehat{\eta}_{n,i}(t)$ defined at the beginning of §4.1. Hence,

$$\mathbb{E}^{\tau}\left[\hat{X}^{e,o}_{n,1}(t,y)\hat{X}^{e,o}_{n,1}(t',y')\right] = \frac{1}{n}\sum_{i=1}^{X^e_n(0,(y-t)^+\wedge(y'-t')^+)} \mathbb{E}^{\tau}\left[\widehat{\eta}_{n,i}(t)\widehat{\eta}_{n,i}(t')\right]$$

$$= \frac{1}{n}\sum_{i=1}^{X^e_n(0,(y-t)^+\wedge(y'-t')^+)} H_{\tau_{n,i}}(t)H^c_{\tau_{n,i}}(t') = \int_0^{(y-t)^+\wedge(y'-t')^+} H_u(t)H^c_u(t')\ d\bar{X}^e_n(0,u). \tag{4.6}$$

Assuming (4.5), which corresponds to convergence of finite measures, from (4.6) we have

$$\mathbb{E}^{\tau}\left[\hat{X}^{e,o}_{n,1}(t,y)\hat{X}^{e,o}_{n,1}(t',y')\right] \to \int_0^{(y-t)^+\wedge(y'-t')^+} H_u(t)H^c_u(t')\ dX_0(u) \equiv K(t_i,y_j,t_{i'},y_{j'}), \tag{4.7}$$

because the integrand is a continuous and bounded real-valued function (see (2.1) of §3.2 in [41]).

That completes this part of the proof, but we also continue to directly show convergence of the covariance functions. Since the random variables in (4.7) are bounded by $\bar{X}_n(0) \leq X^{\uparrow} < \infty$ (applying Assumption 3), we also have convergence of the means associated with the convergence in (4.7), yielding convergence of the covariance functions after unconditioning, i.e.,

$$\begin{aligned} K_n(t,y,t',y') &\equiv \mathbb{E}\Big[\mathbb{E}^{\tau}\Big[\hat{X}^{e,o}_{n,1}(t,y)\hat{X}^{e,o}_{n,1}(t',y')\Big]\Big] \\ &= \mathbb{E}\left[\int_0^{(y-t)^+\wedge(y'-t')^+} H_u(t)H^c_u(t')\ d\bar{X}^e_n(0,u)\right] \\ &\to \int_0^{(y-t)^+\wedge(y'-t')^+} H_u(t)H^c_u(t')\ dX_0(u) \equiv K(t,y,t',y'). \end{aligned} \tag{4.8}$$

As an immediate consequence of (4.8), we have an expression for the variance functions and their convergence,

$$\begin{aligned} \sigma^2_n(t,y) \equiv K_n(t,y,t,y) = \mathbb{E}\left[\int_0^{(y-t)^+} H_u(t)H^c_u(t)\ d\bar{X}^e_n(0,u)\right] \\ \to \int_0^{(y-t)^+} H_u(t)H^c_u(t)\ dX_0(u) \equiv \sigma^2(t,y). \end{aligned} \tag{4.9}$$

4.2.2. *Step 2: Convergence of the fidis in* $\mathbb{D}_{\mathbb{D}}$. We again apply Lemma 4.1 and start assuming (4.5). Hence, for each n, we condition upon the ages.

Step 2a: Joint convergence in $\mathbb{R}^{k\times m}$**.** Fix $m \geq 1$ and $0 < y_1 < \cdots < y_m$. The convergence of the fidis of the vector $\left(\hat{X}^{e,o}_{n,1}(t_1,\cdot),\ldots,\hat{X}^{e,o}_{n,1}(t_k,\cdot)\right)$ in the second argument is equivalent to the joint convergence of the bigger vector $\left(\hat{X}^{e,o}_{n,1}(t_i,y_j), 1\leq i\leq k, 1\leq j\leq m\right)$ in $\mathbb{R}^{k\times m}$. By the Cramér-Wold device (see Theorem 4.3.3 of [41]), this is equivalent to showing that, for all $\{a_{i,j}\} \in \mathbb{R}$, $i=1,\ldots,k$ and $j=1,\ldots,m$, as $n\to\infty$,

$$\sum_{i=1}^{k}\sum_{j=1}^{m} a_{i,j}\hat{X}^{e,o}_{n,1}(t_i,y_j) \Rightarrow \sum_{i=1}^{k}\sum_{j=1}^{m} a_{i,j}\hat{X}^{e,o}_{1}(t_i,y_j) \stackrel{\mathrm{d}}{=} \mathcal{N}(0,\Sigma) \quad \text{in} \quad \mathbb{R}, \tag{4.10}$$

where the variance of the limit is

$$\Sigma \equiv \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{i'=1}^{k}\sum_{j'=1}^{m} a_{i,j}a_{i',j'}K(t_i,y_j,t_{i'},y_{j'}). \tag{4.11}$$

To establish (4.10), we define the random variables

$$\tilde{X}_{n,l,i} \equiv \frac{1}{\sqrt{n}}\widehat{\eta}_{n,l}(t_i) \quad \text{and} \quad \tilde{Y}_{n,l} \equiv \sum_{i=1}^{k}\sum_{j=1}^{m} a_{i,j}\tilde{X}_{n,l,i}\,\mathbf{1}(l \leq X_n(0,(y_j-t_i)^+)).$$

Since $\tilde{Y}_{n,j}$, $j\geq 1$, are independent random variables, conditioned on $X^e_n(0,\cdot)$, we can rewrite the left-hand side of (4.10) as

$$\tilde{S}_n \equiv \sum_{i=1}^{k}\sum_{j=1}^{m} a_{i,j}\hat{X}^{e,o}_{n,1}(t_i,y_j) = \sum_{i=1}^{k}\sum_{j=1}^{m} a_{i,j}\sum_{l=1}^{X^e_n(0,(y_j-t_i)^+)}\tilde{X}_{n,l,i} = \sum_{l=1}^{X^e_n(0,M)}\tilde{Y}_{n,l},$$

where $M \equiv \max\{(y_j - t_i)^+ : 1 \le i \le k, 1 \le j \le m\}$. By the final expression above, $\tilde{S}_n$ is a sum of independent r.v.'s. Of course, the summands $\tilde{Y}_{n,l}$ and the index $X_n^e(0,M)$ both depend on n, but they do so in a regular way because, to apply Lemma 4.1, we are assuming that (4.5) holds. For example, this means that $n^{-1}X_n^e(0,M) \to X^e(0,M) \le X^\uparrow < \infty$.

Hence, we can now apply the Lindeberg-Feller CLT for a double sequence (triangular array) of non-identically distributed independent random variables, e.g., Theorem 7.2.4 of [12]. The variance of $\tilde{S}_n$ is

$$\begin{aligned} \tilde{s}_n^2 \equiv \mathrm{Var}(\tilde{S}_n)) &= \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{i'=1}^{k}\sum_{j'=1}^{m} a_{i,j}a_{i',j'}K_n(t_i,y_j,t_{i'},y_{j'}) \\ &\to \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{i'=1}^{k}\sum_{j'=1}^{m} a_{i,j}a_{i',j'}K(t_i,y_j,t_{i'},y_{j'}) \equiv \Sigma, \end{aligned} \tag{4.12}$$

as $n \to \infty$ where Σ is defined in (4.11) and the convergence follows from (4.8). It remains to verify the Lindeberg conditions (see (2.1)-(2.2) on p.330 of [12]) or the Lyapounov condition (see (2.20) on p.339 of [12]). However, since $\{\tilde{X}_{n,l,i}\}$ take values in the interval $[-1/\sqrt{n}, 1/\sqrt{n}]$ and the variance $\tilde{s}_n$ converges to Σ in (4.12) as $n \to \infty$, the Lindeberg condition is satisfied. Therefore, by Theorem 7.2.4 of [12], if (4.5) holds, then

$$\tilde{S}_n/\tilde{s}_n \Rightarrow \mathcal{N}(0,1) \quad \text{as } n \to \infty, \tag{4.13}$$

which together with (4.12), imply the desired convergence in (4.10) under the condition (4.5). Lemma 4.1 then provides the unconditional convergence.

Step 2b: Tightness in $\mathbb{D}^k$. We now establish the tightness of the vector $(\hat{X}_{n,1}^{e,o}(t_1,\cdot),\ldots,\hat{X}_{n,1}^{e,o}(t_k,\cdot))$ in $\mathbb{D}^k$, again assuming (4.5). This tightness is equivalent to the tightness of each component $\hat{X}_{n,1}^{e,o}(t_i,\cdot)$ in $\mathbb{D}$, for all $1 \le i \le k$, by Theorem 11.6.7. of [41].

We prove tightness of the components by proving a stronger result. Paralleling §7 of [23], we show that, for each fixed t, as $n \to \infty$,

$$\hat{X}_{n,1}^{e,o}(t,y) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n\bar{X}_n^e(0,y-t)} \widehat{\eta}_{n,i}(t) \Rightarrow \mathcal{B}(\sigma_t^2(y)) \quad \text{in} \quad \mathbb{D}, \tag{4.14}$$

where $\mathcal{B}$ is a standard BM. To prove this, we observe that, under the conditioning on the ages, the process $\{\hat{X}_{n,1}^{e,o}(t,y) : y \ge 0\}$ with fixed t is a martingale with respect to its natural filtration augmented by the age sequence $\{\tau_{n,i}\}$ since the function $y \mapsto \bar{X}_n^e(0, y-t)$ is strictly increasing for each $n \ge 1$. Then we can apply the martingale FCLT in Theorem 7.1.4 of

[6], exploiting the fact that the summands are independent $[-1,1]$-valued zero-mean random variables. The first variance function in (4.9) with t fixed is the quadratic variation of $\{\hat{X}^{e,o}_{n,1}(t,y) : y \geq 0\}$ and converges to the second variance formula in (4.9). Hence the function $\sigma^2_t(y)$ in (4.14) for each fixed t. Note that the weak convergence of the components implies their tightness.

4.2.3. *Step 3: Proof of C-tightness of $\{\hat{X}^{e,o}_{n,1}\}$ in $\mathbb{D}_{\mathbb{D}}$.* To complete the proof of the convergence of the first term of (4.3) under condition (4.5), we next show that the sequence $\{\hat{X}^{e,o}_{n,1}\}$ is C-tight in $\mathbb{D}_{\mathbb{D}}$. To do so, we verify the usual two conditions: (i) stochastic boundedness and (ii) asymptotically negligible oscillations, as in Theorem 6.2 of [30]. The specific conditions we establish are (4.15) and (4.26) below.

Verifying condition (i): Stochastic Boundedness. Let $\mathbb{P}^\tau$ and $\mathbb{E}^\tau$ be the conditional probability and expectation given the ages $\{\tau_{n,i}\}$. It suffices to show, under condition (4.5), that for all $\epsilon > 0$, there exists $c > 0$ such that

$$\mathbb{P}^\tau(\|\hat{X}^{e,o}_{n,1}\|_{T,y^\uparrow} > c) \leq \epsilon \quad \text{for all} \quad n \geq 1, \tag{4.15}$$

where $\|\hat{X}^{e,o}_{n,1}\|_{T,y^\uparrow} = \sup_{(t,y)\in[0,T]\times[0,y^\uparrow]} |\hat{X}^{e,o}_{n,1}(t,y)|$.

To bound the probability in (4.15), we apply Chernoff's inequality (e.g., see Lemma 3.1.1. of [12]), obtaining, for $r > 0$,

$$\mathbb{P}^\tau(\|\hat{X}^{e,o}_{n,1}\|_{T,y^\uparrow} > c) \leq e^{-rc}\, \mathbb{E}^\tau\left[e^{r\|\hat{X}^{e,o}_{n,1}\|_{T,y^\uparrow}}\right], \tag{4.16}$$

To bound the right side of (4.16), we follow the symmetrization argument used in the proof of inequality 3 on p.820 of [36] (which in turn follows Lemma 1.1 of [27]). Let the sequence $\{\eta^*_i(\tau_{n,i})\}$ be an independent copy of $\{\eta_i(\tau_{n,i})\}$ conditional on $\{\tau_{n,i}\}$ and let ξ_i be i.i.d. random variables, independent of $\{\eta_i(\tau_{n,i})\}$, with $P(\xi_i = 1) = P(\xi_i = -1) = 1/2$. Also let $\hat{X}^{*,e,o}_{n,1}$ be $\hat{X}^{e,o}_{n,1}$ with $\{\eta_i(\tau_{n,i})\}$ replaced by $\{\eta^*_i(\tau_{n,i})\}$. Let $\mathbb{E}^\tau_\eta \mathbb{E}^\tau_{\eta^*} \mathbb{E}^\tau_\xi$ denote the expectation with respect to $\{\xi_i\}$ conditioned on $\{\tau_{n,i}\}$, $\{\eta_i\}$ and $\{\eta^*_i\}$; let $\mathbb{E}^\tau_\eta \mathbb{E}^\tau_{\eta^*}$ denote the expectation with respect to $\{\eta^*_{n,i}\}$ conditioned on $\{\tau_{n,i}\}$ and $\{\eta_i\}$; and so on.

The next two lemmas will be used in the proof.

LEMMA 4.2. *Let X and X^* be two i.i.d. elements in space $\mathbb{D}_{\mathbb{D}}$. Suppose a function ϕ is convex and nondecreasing with domain $[0,\infty)$. Then*

$$\mathbb{E}\left[\phi(\|X - \mathbb{E}[X]\|_{T,y^\uparrow})\right] \leq \mathbb{E}\left[\phi(\|X - X^*\|_{T,y^\uparrow})\right]. \tag{4.17}$$

We omit the proof because it is similar to that of Lemma (A.14.15) in [36], but we do give the proof in the appendix of the longer online version.

LEMMA 4.3. *Let ζ_i and w_i be real-valued numbers with $w_i \geq 0$ for $1 \leq i \leq N$ and let $\{w_{(i)}, 1 \leq i \leq N\}$ be the order statistics of $\{w_i, 1 \leq i \leq N\}$ with $w_{(i)} \leq w_{(i+1)}$ for $1 \leq i \leq N-1$. For $T > 0$,*

$$\left|\sum_{i=1}^{N} \zeta_i \mathbf{1}(w_{(N-i+1)} > t)\right| \leq \max_{1\leq j\leq N} \left|\sum_{i=1}^{j} \zeta_i\right|, \quad 0 \leq t \leq T. \tag{4.18}$$

Proof. We partition the interval $[0, T]$ into disjoint intervals with end points $0 \equiv w_{(0)} \leq w_{(1)} \wedge T \leq \cdots \leq w_{(N)} \wedge T \leq w_{(N+1)} \equiv T$. We have

$$\begin{aligned}
\left|\sum_{i=1}^{N} \zeta_i\, \mathbf{1}(w_{(N-i+1)} > t)\right| &= \left|\sum_{j=1}^{N} \left(\sum_{i=1}^{j} \zeta_i\right) \mathbf{1}(w_{(N-j)} \wedge T \leq t < w_{(N-j+1)} \wedge T)\right| \\
&\leq \sum_{j=1}^{N} \left|\sum_{i=1}^{j} \zeta_i\right| \mathbf{1}(w_{(N-j)} \wedge T \leq t < w_{(N-j+1)} \wedge T) \\
&\leq \max_{1\leq j\leq N} \left|\sum_{i=1}^{j} \zeta_i\right|. \quad \square
\end{aligned}$$

We now continue to bound the right side of (4.16). For that purpose, define

$$\begin{aligned}
\Gamma_{n,i}(t, y) &\equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{X_n^e(0,(y-t)^+)} \xi_i \cdot \mathbf{1}(\eta_i(\tau_{n,i}) > t) \quad \text{and} \\
\Gamma_{n,i}^*(t, y) &\equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{X_n^e(0,(y-t)^+)} \xi_i \cdot \mathbf{1}(\eta_i^*(\tau_{n,i}) > t).
\end{aligned} \tag{4.19}$$

Let $N_n^*(y, t) \equiv X_n^e(0, (y - t)^+)$ and $\{\bar{\eta}_{n,(i)}, 1 \leq i \leq N_n^*(y, t)\}$ be the order statistics of $\{\eta_i(\tau_{n,i}), 1 \leq i \leq N_n^*(y, t)\}$ so that $\bar{\eta}_{n,(N_n^*(y,t)-i+1)}$ is the i^{th} largest one.

With that preparation, we can write (explanation given afterwards)

$$\begin{aligned}
&\mathbb{E}^\tau\left[\exp\left(r\, \|\hat{X}_{n,1}^{e,o}\|_{T,y^\uparrow}\right)\right] \leq \mathbb{E}^\tau\left[\exp\left(r\, \|\hat{X}_{n,1}^{e,o} - \hat{X}_{n,1}^{*,e,o}\|_{T,y^\uparrow}\right)\right] \\
&= \mathbb{E}^\tau\left[\exp\left(r\, \left\|n^{-1/2} \sum_{i=1}^{N_n^*(y,t)} (\mathbf{1}(\eta_i(\tau_{n,i}) > t) - \mathbf{1}(\eta_i^*(\tau_{n,i}) > t))\right\|_{T,y^\uparrow}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}^{\tau}\left[\exp\left(r\left\|n^{-1/2}\sum_{i=1}^{N_n^*(y,t)}\xi_i\cdot\left(\mathbf{1}(\eta_i(\tau_{n,i})>t)-\mathbf{1}(\eta_i^*(\tau_{n,i})>t)\right)\right\|_{T,y^\uparrow}\right)\right]\\
&\le \mathbb{E}^{\tau}\left[\exp\left(r\ \|\Gamma_{n,i}(t,y)\|_{T,y^\uparrow}+r\ \left\|\Gamma^*_{n,i}(t,y)\right\|_{T,y^\uparrow}\right)\right]\\
&= \mathbb{E}^{\tau}_{\eta}\mathbb{E}^{\tau}_{\eta^*}\mathbb{E}^{\tau}_{\xi}\left[\exp\left(r\,\|\Gamma_{n,i}(t,y)\|_{T,y^\uparrow}+r\,\left\|\Gamma^*_{n,i}(t,y)\right\|_{T,y^\uparrow}\right)\right]\\
&\le \mathbb{E}^{\tau}_{\eta}\left[\mathbb{E}^{\tau}_{\xi}\exp\left(2r\ \|\Gamma_{n,i}(t,y)\|_{T,y^\uparrow}\right)\right]^{1/2}\cdot\mathbb{E}^{\tau}_{\eta^*}\left[\mathbb{E}^{\tau}_{\xi}\exp\left(2r\ \left\|\Gamma^*_{n,i}(t,y)\right\|_{T,y^\uparrow}\right)\right]^{1/2}\\
&= \mathbb{E}^{\tau}_{\eta}\mathbb{E}^{\tau}_{\xi}\left[\exp\left(2r\ \|\Gamma_{n,i}(t,y)\|_{T,y^\uparrow}\right)\right]\\
&= \mathbb{E}^{\tau}_{\eta}\mathbb{E}^{\tau}_{\xi}\left[\exp\left(2r\left\|n^{-1/2}\sum_{i=1}^{N_n^*(y,t)}\xi_i\cdot\mathbf{1}(\bar{\eta}_{n,(N_n^*(y,t)-i+1)}>t)\right\|_{T,y^\uparrow}\right)\right]\\
&\le \mathbb{E}^{\tau}_{\xi}\left[\exp\left(2r\sup_{(t,y)\in[0,T]\times[0,y^\uparrow]}\left\{\max_{1\le j\le N_n^*(y,t)}\left|n^{-1/2}\sum_{i=1}^{j}\xi_i\right|\right\}\right)\right]
\end{aligned}
$$

(4.20)

$$
\le \mathbb{E}^{\tau}_{\xi}\left[\exp\left(2r\max_{1\le j\le X_n(0)}\left|n^{-1/2}\sum_{i=1}^{j}\xi_i\right|\right)\right],
$$

where the *first inequality* holds by Lemma 4.2; the *first equality* holds because the centering terms cancel out; the *second equality* holds because $\xi_i\cdot(\mathbf{1}(\eta_i(\tau_{n,i})>t)-\mathbf{1}(\eta_i^*(\tau_{n,i})>t))\overset{\mathrm{d}}{=}\mathbf{1}(\eta_i(\tau_{n,i})>t)-\mathbf{1}(\eta_i^*(\tau_{n,i})>t)$; the *second inequality* follows by (4.19) and the triangle inequality; the *third* equality holds by conditioning on the η and η^*; the *third inequality* holds by applying the Cauchy-Schwarz inequality on the expectation $\mathbb{E}^{\tau}_{\xi}$; the *fourth equality* holds because $\{\eta_i(\tau_{n,i})\}$ and $\{\eta_i^*(\tau_{n,i})\}$ are two i.i.d. copies; the *fifth equality* holds because the two sequences $\{\xi_i\}$ and $\{\eta_i(\tau_{n,i})\}$ are independent; the *fourth inequality* holds by Lemma 4.3; and the *last inequality* holds because t and y appear only in the upper limit of the inner maximum $N_n^*(y,t)$, which itself is bounded above by $X_n(0)$.

To bound (4.20), we apply integration by parts as on p. 150 of [7] to write the moment generating function of a non-negative random variable Y as

$$
\mathbb{E}\left[e^{\theta Y}\right]=1+\int_0^\infty\theta e^{\theta x}\mathbb{P}\left(Y\ge x\right)\ dx. \tag{4.21}
$$

Therefore we next provide an upper bound on the tail probability using Lévy's inequality .(e.g., Theorem 3.7.1 on p.138 of [12] (also Theorem B.1

in the appendix of the longer online version); in particular

(4.22)

$$\mathbb{P}^\tau\left(\max_{1\le j\le X_n(0)}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{j}\xi_i\right|\ge x\right)\le 2\mathbb{P}^\tau\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{X_n(0)}\xi_i\right|\ge x\right)\le 4e^{-\frac{x^2}{2\bar{X}_n(0)}},$$

where the *second* inequality follows from Hoeffding's inequality (e.g., Theorem 3.1.3 on p.120 of [12]; also Theorem B.2 in the appendix).

Combining (4.20)–(4.22) with $\theta = 2r$ and $Y \equiv \max_{1\le j\le X_n(0)}\left|1/\sqrt{n}\sum_{i=1}^{j}\xi_i\right|$ yields that

$$\mathbb{E}^\tau\left[\exp\left(r\|\hat{X}^{e,o}_{n,1}\|_{T,y^\uparrow}\right)\right]\le 1+4\int_0^\infty 2re^{2rx}e^{-\frac{x^2}{2\bar{X}_n(0)}}\,dx$$

$$=1+8r\sqrt{2\pi\bar{X}_n(0)}e^{2r^2\bar{X}_n(0)}\int_0^\infty\left(2\pi\bar{X}_n(0)\right)^{-1/2}e^{-\frac{(x-2r\bar{X}_n(0))^2}{2\bar{X}_n(0)}}\,dx$$

(4.23)

$$=1+8r\sqrt{2\pi\bar{X}_n(0)}e^{2r^2\bar{X}_n(0)}\Phi\left(2r\sqrt{\bar{X}_n(0)}\right)\le 1+8r\sqrt{2\pi\bar{X}_n(0)}e^{2r^2\bar{X}_n(0)},$$

where Φ is the cdf of the standard normal distribution. Here we assume without loss of generality that $\bar{X}_n(0) > 0$ because (4.20) becomes 0 if $\bar{X}_n(0) = 0$.

Now recall that we are assuming (4.5), so that we have $\bar{X}_n(0)\to X(0)$ w.p.1 as $n\to\infty$. Moreover, this convergence implies that there exists a constant K such that

$$\bar{X}_n(0)\le 2X(0)+K\equiv X^\uparrow\quad\text{for all}\quad n\ge 1, \tag{4.24}$$

where $X^\uparrow$ depends on the particular age sequence associated with our conditioning.

Letting $r = 1/\sqrt{\bar{X}_n(0)}$ in (4.23) and applying (4.16), we have

$$\begin{aligned}\mathbb{P}^\tau(\|\hat{X}^{e,o}_{n,1}\|_{T,y^\uparrow}>c)&\le\left(1+8\sqrt{2\pi}e^2\right)\mathbb{E}^\tau\left[\exp\left(-\frac{c}{\sqrt{\bar{X}_n(0)}}\right)\right]\\&\le\left(1+8\sqrt{2\pi}e^2\right)\exp\left(-\frac{c}{\sqrt{X^\uparrow}}\right),\end{aligned}$$

which converges to 0 as $c\to\infty$.

Verifying condition (ii): asymptotically negligible oscillations. We show that the oscillations are asymptotically negligible, again assuming (4.5). For that purpose, consider an arbitrary sequence of uniformly bounded stopping times $\{\kappa_n\}$ with respect to the natural filtration $\mathbf{F}_n \equiv \{\mathcal{F}_n(t), t \in [0,\infty)\} \vee \mathcal{N}$ where

$$\begin{aligned} \mathcal{F}_n(t,y) &\equiv \sigma\{1(\eta_i(\tau_{n,i}) > x) : 1 \le i \le X_n^e(0,y), x \ge t, 0 \le s \le y\} \\ &\quad \vee \sigma\{X_n^e(0,(y-x)^+), x \ge t, 0 \le s \le y\}, \\ \mathcal{F}_n(t) &\equiv \bigvee_{y \ge 0} \mathcal{F}_n(t,y), \end{aligned} \tag{4.25}$$

and $\mathcal{N}$ being all the null sets. We will show that, for any $\delta > 0$ and $\epsilon > 0$, and for any such sequence of stopping times $\{\kappa_n\}$,

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} \sup_{\kappa_n} \mathbb{E}^\tau \left[\left(\sup_{y \ge 0} \left| \hat{X}^{e,o}_{n,1}(\kappa_n + \delta, y) - \hat{X}^{e,o}_{n,1}(\kappa_n, y) \right| \right)^2 \right] = 0, \tag{4.26}$$

which is a sufficient condition for condition (ii) in Theorem 6.2 of [30].

To establish (4.26), we condition on the sequence $\{\kappa_n\}$ as well as the sequence $\{\tau_{n,i}\}$. As in Step 2b, conditional on the sequences $\{\kappa_n\}$ and $\{\tau_{n,i}\}$, the process $\{\hat{X}^e_{n,1}(t,y) : y \ge 0\}$ with t fixed is an adapted martingale with respect to $\tilde{\mathbf{F}}^t_n \equiv \bigvee_{y \ge 0} \mathcal{F}^t_n(y) \vee \{\kappa_n\} \vee \{\tau_{n,i}\}$, where $\mathcal{F}^t_n(y)$ denotes the σ-algebra $\mathcal{F}_n(t,y)$ in (4.25) with t being fixed. Consequently, with the conditioning, the process $(\hat{X}^{e,o}_{n,1}(\kappa_n + \delta, y) - \hat{X}^{e,o}_{n,1}(\kappa_n, y),\ y \ge 0)$ is an $\tilde{\mathbf{F}}^{\kappa_n+\delta}_n$-adapted martingale. Let $\mathbb{E}^{\tau,\kappa}$ denote that the expectation is computed by conditioning on $\{\tau_{n,i}\}$ and $\{\kappa_n\}$, let $\mathrm{Var}^{\tau,\kappa}$ be the conditional variance. Then, by Doob's maximal inequality,

$$\begin{aligned} &\mathbb{E}^{\tau,\kappa} \left(\sup_{y \ge 0} \left| \hat{X}^{e,o}_{n,1}(\kappa_n + \delta, y) - \hat{X}^{e,o}_{n,1}(\kappa_n, y) \right| \right)^2 \\ &\le 4 \sup_{y \ge 0} \mathbb{E}^{\tau,\kappa} \left(\hat{X}^{e,o}_{n,1}(\kappa_n + \delta, y) - \hat{X}^{e,o}_{n,1}(\kappa_n, y) \right)^2 \\ &= \frac{4}{n} \sup_{y \ge 0} \ \mathrm{Var}^{\tau,\kappa} \left(\sum_{i=1}^{X_n^e(0,(y-\kappa_n-\delta)^+)} \mathbf{1}(\eta_i(\tau_{n,i}) > \kappa_n + \delta) - \sum_{i=1}^{X_n^e(0,(y-\kappa_n)^+)} \mathbf{1}(\eta_i(\tau_{n,i}) > \kappa_n) \right) \\ &\le \frac{4}{n} \sup_{y \ge 0} \ \mathrm{Var}^{\tau,\kappa} \left(\sum_{i=1}^{X_n^e(0,(y-\kappa_n)^+)} \mathbf{1}(\kappa_n < \eta_i(\tau_{n,i}) \le \kappa_n + \delta) \right) \\ &= \frac{4}{n} \sup_{y \ge 0} \sum_{i=1}^{X_n^e(0,(y-\kappa_n)^+)} H^\delta_{\tau_{n,i}}(\kappa_n)(1 - H^\delta_{\tau_{n,i}}(\kappa_n)) \\ &= 4 \int_0^\infty H^\delta_u(\kappa_n)(1 - H^\delta_u(\kappa_n)) \, d\bar{X}^e_n(0,u) \end{aligned}$$

$$\leq 4\int_0^\infty H_u^\delta(M)\ d\bar{X}_n^e(0,u) \leq \int_0^\infty \frac{4g_0^\uparrow\delta}{G^c(T+y^\uparrow)}\ d\bar{X}_n^e(0,u)$$

(4.27)

$$= \frac{4g_0^\uparrow\delta}{G^c(T+y^\uparrow)}\bar{X}_n(0) \leq \frac{4g_0^\uparrow\delta}{G^c(T+y^\uparrow)}X^\uparrow \to 0,$$

as $\delta \to 0$. Therefore, the condition in (4.26) is satisified. In the steps above, $H_u^\delta(t) \equiv H_u(t+\delta) - H_u(t)$, $M = \sup_{n\geq 1}|\kappa_n|$ and $X^\uparrow$ is the bound in (4.24). The first equality holds since the sums are zero-mean random variables conditioned on $\{\tau_{n,i}\}$, $\{\kappa_n\}$ for all $t \geq 0$, $y \geq 0$, whereas the second equality holds due to (conditional) independence. Starting from the third equality, $\{\tau_{n,i}\}$ and $\{\kappa_n\}$ are necessarily treated as deterministic sequences. Assumption 2 implies that we are not dividing by 0 in the final step.

4.3. *Convergence of the Second Term in* (4.3). In this section, we establish convergence of the second term in (4.3), i.e., $\hat{X}_{n,2}^{e,o} \Rightarrow \hat{X}_2^{e,o}$, again conditioning on the ages and assuming that (4.5) holds, so that we can apply Lemma 4.1. Since $\hat{X}_n^e(0,\cdot)$ is of bounded variation, the second term in (4.3) can be expressed as a Stieltjes integral. Therefore, we can use the integration by parts formula given on p.336 of [2] to obtain an equivalent representation

$$\text{(4.28)}\quad \hat{X}_{n,2}^{e,o}(t,y) \equiv -\int_0^{(y-t)^+} H_u^c(t)d\hat{X}_n^e(0,u)$$

$$= H_{(y-t)^+}^c(t)\hat{X}_n^e(0,(y-t)^+) - \int_0^{(y-t)^+} \hat{X}_n^e(0,u-)dH_u^c(t).$$

Since we are conditioning on the ages, everything in (4.28) is deterministic. Hence, we will show that the convergence follows by continuity (convergence preservation of mappings). The mapping is a measurable mapping that is continuous almost surely with respect to continuous limits. Measurability in this setting holds because the Borel σ-field induced by the usual topology on $\mathbb{D}_\mathbb{D}$ coincides with the usual Kolmogorov σ-field generated by the coordinate projections; see §11.5.3 of [41] and references cited there. (Hence standard measurability arguments can be used.) If we uncondition, then we would be applying the continuous mapping theorem in Theorem 3.4.3 of [41].

The next two lemmas allow us to establish the desired convergence. We first show that the function $H_x^c(t)$ has finite variation in x over a bounded interval, by virtue of the Assumption 2 on the service-time cdf G.

LEMMA 4.4 (Finite total variation in x for $H^c_x(t)$ in bounded intervals). *In an interval* $[0, T^*]$, $\int_0^{T^*} |dH^c_x(t)| < \infty$, *for* $t \geq 0$.

Proof. Taking the derivative of $H^c_x(t)$ with respect to x yields

$$\begin{aligned} \int_0^{T^*} |dH^c_x(t)| &= \int_0^{T^*} \frac{|g(t+x)G^c(x) - g(x)G^c(t+x)|}{(G^c(x))^2} dx \\ &< \left(\frac{g^\uparrow}{G^c(T^*)} + \frac{g^\uparrow}{(G^c(T^*))^2} \right) T^* \equiv K(T^*) < \infty, \end{aligned} \tag{4.29}$$

where we have used Assumption 2. □

We next establish continuity in the uniform metric over compact subsets of the domain. Let $d_u(x_1, x_2) \equiv \sup_{t \in [0,T]} |x_1(t) - x_2(t)|$ for $x_1, x_2 \in \mathbb{D}$ and

$$d_u(y_1, y_2) \equiv \sup_{(t,u) \in [0,T] \times [0,\infty)} |y_1(t,u) - y_2(t,u)| \quad \text{for} \quad y_1, y_2 \in \mathbb{D}_{\mathbb{D}}.$$

LEMMA 4.5. *The mapping* $\phi : (\mathbb{D}, d_u) \to (\mathbb{D}_{\mathbb{D}}, d_u)$ *defined by*

$$\phi(x)(t,y) = H_{(y-t)^+}(t) x((y-t)^+) - \int_0^{(y-t)^+} x(s-) dH_s(t) \tag{4.30}$$

for $0 \leq t \leq y$ *is continuous in* $\mathbb{D}_{\mathbb{D}}$.

Proof. Let $\{x_n\}$ be a sequence such that $d_u(x_n, x) \equiv \|x_n - x\|_{y_0} \to 0$ as $n \to \infty$. Then

$$\begin{aligned} \left|\phi(x_n)(t,y) - \phi(x)(t,y)\right| &\leq H^c_{(y-t)^+}(t) \left|x_n((y-t)^+) - x((y-t)^+)\right| \\ &\quad + \left| \int_0^{(y-t)^+} (x_n(s-) - x(s-))\, dH^c_s(t) \right| \\ &\leq H^c_{(y-t)^+}(t) \|x_n - x\|_{y_0} + \|x_n - x\|_{y_0} \int_0^{(y-t)^+} |dH^c_s(t)| \\ &\leq (1 + K(y_0)) \|x_n - x\|_{y_0}, \end{aligned}$$

where the finite constant $K(y_0)$ is defined in (4.29). Therefore, as $n \to \infty$,

$$d_u(\phi(x_n), \phi(x)) \equiv \sup_{(t,y) \in [0,T] \times [0,\infty)} |\phi(x_n)(t,y) - \phi(x)(t,y)| \to 0. \quad \square \tag{4.31}$$

Finally, we observe that Lemma 4.5 establishes the desired result, because (i) it suffices to consider continuous limits x by virtue of the continuity assumption included in Assumption 1 and (ii) convergence in $\mathbb{D}$ reduces to uniform convergence over bounded intervals when the limit function is continuous.

4.4. *Proof of convergence of other processes.* The proof of convergence of the other processes is elementary. First, we can apply flow conservation and continuous mapping theorem to treat the departure process. In particular, as $n \to \infty$,

$$\hat{D}_n(t) = \hat{N}_n(t) + \hat{X}_n(0) - \hat{X}_n(t) \Rightarrow \hat{N}(t) + \hat{X}(0) - \hat{X}(t),$$

which coincides with (3.14).

To treat the remaining-service-time processes $\hat{X}_n^r(0,x)$ and $\hat{X}_n^r(t,x)$, note that $X_n^r(0,x) = X_n^o(x) = X_n^o(x,\infty)$ which is the number of initially existing customers that are still in service at time x. Hence, as $n \to \infty$, $\hat{X}_n^r(0,x) = \hat{X}_n^o(x) \Rightarrow \hat{X}^o(x)$ in $\mathbb{D}$, which is proved earlier in this section. Next, just as for the ICP $\hat{X}_n^e(t,x)$, we split $\hat{X}_n^r(t,x)$ into two independent terms associated with new content and old content,

$$\hat{X}_n^r(t,x) = \hat{X}_n^{r,\nu}(t,x) + \hat{X}_n^{r,o}(t,x),$$

where the convergence of $\hat{X}_n^{r,\nu}(t,x)$ is proved in [30] and the convergence of the second term holds because

$$\hat{X}_n^{r,o}(t,x) = \hat{X}_n^r(0,t+x) = \hat{X}_n^o(t+x) \Rightarrow \hat{X}^o(t+x), \quad \text{as} \quad n \to \infty, \quad \text{in} \quad \mathbb{D}.$$

4.5. *Proof of alternative representations in* (3.11) *and* (3.16). We only prove (3.11) because (3.16) is similar. We obtain the right-hand representation in (3.11) using the fact that $\{\hat{U}(t,y_0)/\sqrt{y_0(1-y_0)}; t \ge 0\}$ is a standard BM motion for a fixed $0 < y_0 < 1$ (§A of the appendix of [30]). We have

$$\begin{aligned}
&\int_0^t \int_0^\infty 1(x > t-s)\, d\hat{K}_\nu(\Lambda(s),x) \\
&= \int_0^t \int_0^\infty 1(x > t-s)\, d\hat{U}(\Lambda(s),G_\nu(x)) \\
&\stackrel{\mathrm{d}}{=}_t \int_0^t \int_0^\infty 1(x > t-s)\, d\left(\tilde{\mathcal{B}}_s(\Lambda(s))\sqrt{G_\nu(x)(1-G_\nu(x))}\right) \\
&= \int_0^t \left(\sqrt{G_\nu(\infty)G_\nu^c(\infty)} - \sqrt{G_\nu(t-s)G_\nu^c(t-s)}\right) d\tilde{\mathcal{B}}_s(\Lambda(s)),
\end{aligned}$$

which coincides with the right-hand expression in (3.11). To show that the two expressions in (3.11) are indeed equal in distribution for each t, it suffices to show that they have the same variances because both processes are zero-mean Gaussian processes. Because the Kiefer process $\hat{U}(\Lambda(s),G_\nu(x)) = \hat{W}(\Lambda(s),G_\nu(x)) - G_\nu(x)\hat{W}(\Lambda(s),1)$ where $\hat{W}$ is a standard Brownian sheet

[30], the variance of the first expression in (3.11) is

$$\mathbb{E}\left[\left(\int_0^t \int_0^\infty \mathbf{1}(x > t-s)\, d\left(\hat{W}(\Lambda(s), G_\nu(x)) - G_\nu(x)\hat{W}(\Lambda(s),1)\right)\right)^2\right]$$
$$= \mathbb{E}\left[\left(\int_0^t \int_0^\infty \mathbf{1}(x > t-s)\, d\hat{W}(\Lambda(s), G_\nu(x)) - \int_0^t G_\nu^c(t-s)\, d\hat{W}(\Lambda(s),1)\right)^2\right]$$
$$= \int_0^t \int_0^\infty \mathbf{1}(x > t-s)\, d\Lambda(s) dG_\nu(x) + \int_0^t G_\nu^c(t-s)^2\, d\Lambda(s)$$
$$- 2\int_0^t \int_0^\infty G_\nu^c(t-s)\mathbf{1}(x > t-s)\, d\Lambda(s) dG_\nu(x)$$
$$= \int_0^t G_\nu^c(t-s) G_\nu(t-s)\, d\Lambda(s),$$

which simply coincides with the variance of the second expression in (3.11).

5. The $G_t/GI^o, GI^\nu/\infty$ model starting in the past.

We now show that Theorem 3.2 applies to the $G_t/GI^o, GI^\nu/\infty$ model starting at some time in the past, provided we impose an extra condition. We assume that the system starts at time $-t_0 < 0$, satisfying the assumptions in §2 with service-time cdf G. We let the service-time cdf change to G_ν after time 0. It suffices to show that Assumption 1 holds at time 0, which requires an additional independent-increments assumption on the arrival process to obtain the assumed independence of the processes. In particular, we assume that the limit process in the assumed FCLT for the arrival process is a time-transformed BM.

Corollary 5.1 (FCLT for the $G_t/GI^o, GI^\nu/\infty$ model starting in the past). *Consider the sequence of $G_t/GI^o, GI^\nu/\infty$ models starting at time $-t_0 < 0$ with all the assumptions in §2 at time $-t_0$. Let the service-time cdf change from G to G_ν at time 0. If in addition $\hat{N}(t) = c_a \mathcal{B}_a(\Lambda(t))$, where $\Lambda(t) \equiv \int_{-t_0}^t \lambda(s)\, ds$, $t > -t_0$, then Assumption 1 also holds at time 0, so that Theorem 3.2 holds for $t \geq 0$, with*

$$X^e(0,y) = \int_0^y G^c(s)\lambda(-s)\, ds \cdot \mathbf{1}(0 \leq y \leq t_0) + \int_0^{y-t_0} H_x^c(t_0 + y)\, dX^e(-t_0, x) \cdot \mathbf{1}(y > t_0), \tag{5.1}$$

$$\hat{X}^e(0,y) = \hat{X}_a^e(0,y) \cdot \mathbf{1}(0 \leq y \leq t_0) + \left(\hat{X}_{b,1}^e(0,y) + \hat{X}_{b,2}^e(0,y)\right) \cdot \mathbf{1}(y > t_0), \tag{5.2}$$

where, with $\mathcal{B}_a$, $\mathcal{B}_s$ *and* $\mathcal{B}$ *denoting three independent standard BM's,* $\hat{K}_o(\Lambda(s),x) \equiv \hat{U}_o(\Lambda(s),G(x))$ *and* $\hat{U}_o$ *denoting a Kiefer process that is independent with* $\mathcal{B}_a$,

$$\begin{aligned}\hat{X}_a^e(0,y) &= c_\lambda \int_{-y}^0 G^c(-s)\,d\mathcal{B}_a(\Lambda(s)) + \int_{-y}^0\int_0^\infty \mathbf{1}(x>-s)\,d\hat{K}_o(\Lambda(s),x)\\ &\stackrel{\mathrm{d}}{=}_y c_\lambda \int_{-y}^0 G^c(-s)\,d\mathcal{B}_a(\Lambda(s)) - \int_{-y}^0 \sqrt{G(-s)G^c(-s)}\,d\mathcal{B}_s(\Lambda(s))\\ &\stackrel{\mathrm{d}}{=}_y \int_{-y}^0 \sqrt{(c_\lambda^2-1)G^c(-s^2)+G^c(-s)\lambda(s)}\,d\mathcal{B}(s), \quad \textit{for} \quad 0\le y\le t_0, \end{aligned} \tag{5.3}$$

$\hat{X}_{b,1}^e(0,y)$ *is a zero-mean Gaussian process with covariance*

$$\mathrm{Cov}\left(\hat{X}_{b,1}^e(0,y_1), \hat{X}_{b,1}^e(0,y_2)\right) = \int_0^{y_1\wedge y_2 - t_0} H_u(t_0)H_u^c(t_0)\,dX^e(-t_0,u),$$

$$\textit{and} \qquad \hat{X}_{b,2}^e(0,y) = \int_0^{y-t_0} H_x^c(t_0)\,d\hat{X}^e(-t_0,u), \quad \textit{for } y, y_1, y_2 > t_0.$$

If the system starts empty at time $-t_0$, *then the variance formula for the FCLT limit of the number in service* $\hat{X}(t)$ *for* $t \ge -t_0$ *is*

$$\begin{aligned}\sigma_{\hat{X}}^2(t) &= \int_{-t_0}^0 \left[(c_\lambda^2-1)G^c(t-s)^2 + G^c(t-s)\right]\lambda(s)\,ds\\ &\quad + \int_0^t \left[(c_\lambda^2-1)G_\nu^c(t-s)^2 + G_\nu^c(t-s)\right]\lambda(s)\,ds.\end{aligned}$$

If in addition $G = G_\nu$, *we have*

$$\sigma_{\hat{X}}^2(t) = \sigma_\nu^2(-t_0,t) = \int_{-t_0}^t \left[(c_\lambda^2-1)G_\nu^c(t-s)^2 + G_\nu^c(t-s)\right]\lambda(s)\,ds. \tag{5.4}$$

REMARK 5.1 (Verifying consistency with [30]). Corollary 5.1 provides an important consistency check by allowing us to compare with the previous results in [30]. In particular, we see that we get strong verification through (5.4).

We illustrate Corollary 5.1 with the following example.

EXAMPLE 5.1 (Simulation comparison). We consider an $M_t/LN(1,4)/\infty$ model over the time interval $[-t_0,T] = [-5,20]$, having a nonhomogeneous Poisson arrival process (M_t) with the sinusoidal arrival rate function

$$\lambda_n(t) = n\left(a + b\sin(ct+\phi)\right), \quad t_0 \le t \le T, \tag{5.5}$$

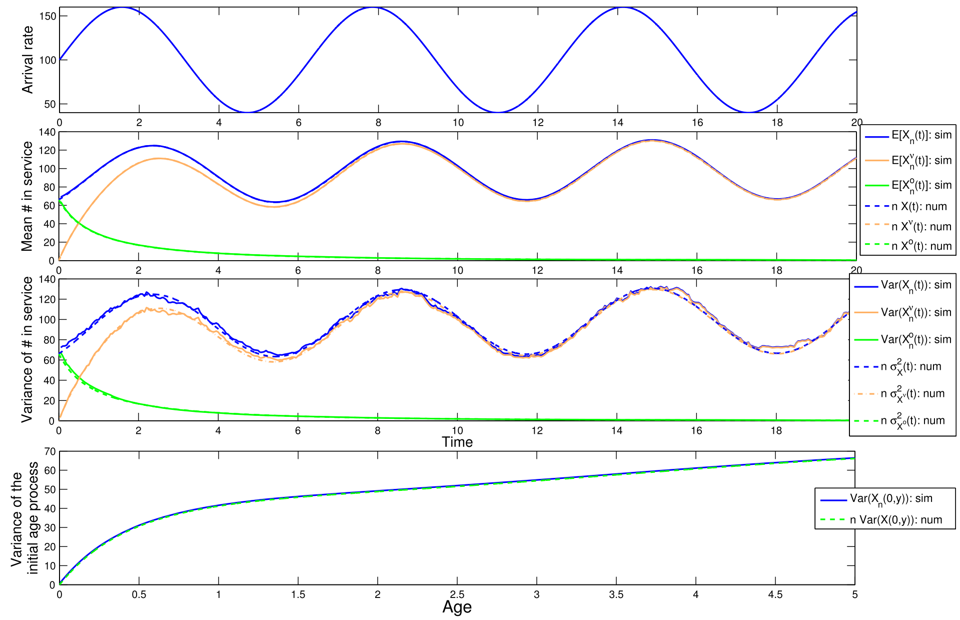

FIG 1. *Example 1: Simulation comparisons of the mean and variance for the number of customers in service in an $M_t/LN(1,4)/\infty$ model starting empty at a finite negative time $-t_0 = -5$, with the sinusoidal arrival rate (5.5) having parameters $a = c = 1$, $b = 0.6$, $\phi = 0$ and $n = 100$.*

with $a = c = 1$, $b = 0.6$ and $\phi = 0$. This example has a lognormal (LN) service distribution with mean $1/\mu = 1$ and $c_s^2 = 4$. We set $n = 100$.

We let the system start empty at time $-t_0$ and use the arrivals in the negative time interval $[-t_0, 0]$ to generate the initial number of customers in service and the age process at time 0. We expect our FWLLN and FCLT limits to provide effective engineering approximations for the mean and variance of the performance functions. For instance, Theorems 3.1 and 3.2 imply that $X_n(t) \approx nX(t) + \sqrt{n}\hat{X}(t)$ when n is large. Therefore, we expect $\mathbb{E}[X_n(t)] \approx nX(t)$ and $\mathrm{Var}(X_n(t)) \approx n\mathrm{Var}(\hat{X}(t))$. We next provide simulation comparison results. Each simulation experiment in this paper is based on performing 2000 independent replications of the system.

Figure 1 shows close approximations of the fluid and variance formulas provided by Theorem 3.1 and Corollaries 3.5 and 5.1. In particular, the arrival rate after time 0 is shown in the top plot. Then the expected number in system of the old customers and the new customers are shown together with the total expected number in the second plot; while the variances of the number in system of the old customers and the new customers are shown together with the total variances in the third plot. In both cases, we see the additivity. As expected, the old content dissipates by about time $t = 6$. The bottom plot shows the variance of the initial age process *at* time 0, with age $0 \le y \le t_0 = 5$ (because no customer has an age greater than t_0). As expected the right endpoint in the bottom plot coincides with the left endpoint in the third plot.

APPENDIX A: TIGHTNESS IN $\mathbb{D}_{\mathbb{D}}$

We now review a necessary and sufficient condition for tightness of a stochastic process $\{X_n : n \ge 1\}$ in space $\mathbb{D}_{\mathbb{D}}$. Also see [30] for details.

LEMMA A.1. *A sequence of stochastic process $\{X_n : n \ge 1\}$ in $\mathbb{D}_{\mathbb{D}}$ is tight if and only if*

(*i*) *The sequence $\{X_n : n \ge 1\}$ is stochastically bounded in $\mathbb{D}_{\mathbb{D}}$, i.e., for all $\epsilon > 0$, there exists a compact subset $K \subset \mathbb{R}$ such that*

$$P(\|X_n\|_T \in K) > 1 - \epsilon, \qquad \text{for all} \quad n \ge 1,$$

where $\|X_n\|_T = \sum_{s\in[0,T]} \sup_{t\in[0,T]} |X_n(s,t)|$; and any one of the following

(*ii*) *For all $\delta > 0$, and all uniformly bounded sequences $\{\tau_n : n \ge 1\}$ where for each n, τ_n is a stopping time with respect to the natural filtration $\mathbf{F}_n = \{\mathcal{F}_n(t), t \in [0,T]\}$ where $\mathcal{F}_n(t) = \sigma\{X_n(s,\cdot) : 0 \le s \le t\}$, there exists a constant $\beta > 0$ such that*

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \sup_{\tau_n} E\left[(1 \wedge d_{J_1}(X_n(\tau_n + \delta, 0), X_n(\tau_n, \cdot)))^\beta\right] = 0;$$

or

(*ii′*) *For all $\delta > 0$, there exists a constant β and random variables $\gamma_n(\delta) \ge 0$ such that for each n, w.p.1.,*

$$E\Big[(1 \wedge d_{J_1}(X_n(s+u,\cdot), X_n(s,\cdot)))^\beta|\mathcal{F}_n\Big](1 \wedge d_{J_1}(X_n(s-v,\cdot), X_n(s,\cdot)))^\beta \le E\left[\gamma_n(\delta)|\mathcal{F}_n\right],$$

for all $0 \le s \le t$, $0 \le u \le \delta$ and $0 \le u \le s \wedge \delta$, where $\mathbf{F}_n = \{\mathcal{F}_n(t) : t \in [0,T]\}$ with $\mathcal{F}_n(t) = \sigma\{X_n(s,\cdot) : 0 \le s \le t\}$ and

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} E\left[\gamma_n(\delta)\right] = 0.$$

APPENDIX B: USEFUL INEQUALITIES

In this section we review two useful inequalities. Both are used to prove (4.22) in §4.

THEOREM B.1 (Lévy's inequality (symmetric case), Theorem 3.7.1 of Gut [12]). *Let $X_1,\ X_2,\dots,X_n$ be independent real-valued symmetric random variables (satisfying $-X_i \overset{\mathrm{d}}{=} X_i$ for all $1 \le i \le n$) and let $S_n = \sum_{k=1}^n X_k$, $n \ge 1$ be the partial sums. Then, for any $x > 0$,*

$$\mathbb{P}\left(\max_{1\le k\le n} |S_k| > x\right) \le 2\mathbb{P}(|S_n| > x). \tag{B.1}$$

THEOREM B.2 (Hoeffding's inequality, Theorem 3.1.3 of Gut [12]). *Let $X_1, X_2,\dots,X_n$ be independent real-valued random variables such that $\mathbb{P}(a_k \le X_k \le b_k) = 1$ for $a_k, b_k \in \mathbb{R}$, $k = 1,\dots,n$, and let $S_n = \sum_{k=1}^n X_k$, $n \ge 1$, denote the partial sums. Then*

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| > x) \le 2\exp\left(-\frac{2x^2}{\sum_{k=1}^n (b_k - a_k)^2}\right). \tag{B.2}$$

APPENDIX C: ADDITIONAL PROOFS

We now provide proofs of Corollaries 3.4 and 5.1, and Lemma 4.2, which were omitted in the main paper.

Proof of Corollary 3.4. This follows from parts (ii) and (iii) of Theorem 3.2. We present the proof of the four-parameter covariance formulas in (i); the variance formulas in (i) and (ii) easily follow.

First, the covariances of $\hat{X}_1^{\nu}$ and $\hat{X}_2^{o}$ are

$$\begin{aligned}
&\mathrm{Cov}(\hat{X}_1^{e,\nu}(t_1, y_1), \hat{X}_1^{e,\nu}(t_2, y_2))\\
&= \mathbb{E}\left[\int_{(t_1-y_1)^+}^{t_1} G_\nu^c(t_1 - s)\, d\hat{N}(s) \times \int_{(t_2-y_2)^+}^{t_2} G_\nu^c(t_2 - s)\, d\hat{N}(s)\right]\\
&= c_\lambda^2 \int_{(t_1-y_1)^+\vee(t_2-y_2)^+}^{t_1\wedge t_2} G_\nu^c(t_1 - s) G_\nu^c(t_1 - s)\, d\Lambda(s),
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Cov}(\hat{X}_2^{e,o}(t_1, y_1), \hat{X}_2^{e,o}(t_2, y_2))\\
&= \mathbb{E}\left[\int_0^{(y_1-t_1)^+} H_x^c(t_1)\, d\hat{X}_0(x) \times \int_0^{(y_2-t_2)^+} H_x^c(t_2)\, d\hat{X}_0(x)\right]\\
&= \int_0^{(y_1-t_1)^+\wedge(y_2-t_2)^+} H_x^c(t_1) H_x^c(t_2)\, d\Sigma_2^{e,o}(x).
\end{aligned}$$

Second, the covariance of $\hat{X}_2^{\nu}$

$$
\begin{aligned}
&\mathrm{Cov}(\hat{X}_2^{e,\nu}(t_1,y_1),\hat{X}_2^{e,\nu}(t_2,y_2))\\
&=\mathbb{E}\left[\left(\int_{(t_1-y_1)^+}^{t_1}\int_0^\infty \mathbf{1}_{(x+s>t_1)}\,d\left(\hat{W}(\Lambda(s),G_\nu(x))-G_\nu(x)\hat{W}(\Lambda(s),1)\right)\right)\right.\\
&\left.\times\left(\int_{(t_2-y_2)^+}^{t_2}\int_0^\infty \mathbf{1}_{(x+s>t_2)}\,d\left(\hat{W}(\Lambda(s),G_\nu(x))-G_\nu(x)\hat{W}(\Lambda(s),1)\right)\right)\right]\\
&=\mathbb{E}\left[\int_{(t_1-y_1)^+}^{t_1}\int_0^\infty \mathbf{1}_{(x+s>t_1)}\,d\hat{W}(\Lambda(s),G_\nu(x))\times\right.\\
&\qquad\left.\int_{(t_2-y_2)^+}^{t_2}\int_0^\infty \mathbf{1}_{(x+s>t_2)}\,d\hat{W}(\Lambda(s),G_\nu(x))\right]\\
&+\mathbb{E}\left[\int_{(t_1-y_1)^+}^{t_1} G_\nu^c(t_1-s)\,d\hat{W}(\Lambda(s),1)\times\int_{(t_2-y_2)^+}^{t_2} G_\nu^c(t_2-s)\,d\hat{W}(\Lambda(s),1)\right]\\
&-\mathbb{E}\left[\int_{(t_1-y_1)^+}^{t_1}\int_0^\infty \mathbf{1}_{(x+s>t_1)}\,d\hat{W}(\Lambda(s),G_\nu(x))\times\int_{(t_2-y_2)^+}^{t_2} G_\nu^c(t_2-s)\,d\hat{W}(\Lambda(s),1)\right]\\
&-\mathbb{E}\left[\int_{(t_1-y_1)^+}^{t_1} G_\nu^c(t_1-s)\,d\hat{W}(\Lambda(s),1)\times\int_{(t_2-y_2)^+}^{t_2}\int_0^\infty \mathbf{1}_{(x+s>t_2)}\,d\hat{W}(\Lambda(s),G_\nu(x))\right]\\
&=\int_{(t_1-y_1)^+\vee(t_2-y_2)^+}^{t_1\wedge t_2} G_\nu^c(t_1\vee t_2-s)d\Lambda(s)+\int_{(t_1-y_1)^+\vee(t_2-y_2)^+}^{t_1\wedge t_2} G_\nu^c(t_1-s)G_\nu^c(t_2-s)\,d\Lambda(s)\\
&-2\int_{(t_1-y_1)^+\vee(t_2-y_2)^+}^{t_1\wedge t_2} G_\nu^c(t_1-s)G_\nu^c(t_2-s)\,d\Lambda(s)\\
&=\int_{(t_1-y_1)^+\vee(t_2-y_2)^+}^{t_1\wedge t_2} G_\nu^c(t_1\vee t_2-s)G_\nu(t_1\wedge t_2-s)\,d\Lambda(s). \quad \square
\end{aligned}
$$

Proof of Corollary 5.1. First, (5.1) and (5.2) easily follow Theorems 3.1 and 3.2 of [30] by considering the performance of a system at the end of interval $[0,t_0]$ (that is, at time t_0) with the system being initially empty (at time 0). Then, it suffices to apply a time shift, that is, shifting the interval to the left by t_0 so that the interval becomes $[-t_0,0]$.

When the system starts empty at $-t_0$, following (5.2) and Theorem 4.2 of [30], the variance function of $\hat{X}^e(0,y)$ is

(C.1)
$$
\mathrm{Var}(\hat{X}^e(0,y))=\int_0^y \left[(c_\lambda^2-1)G^c(s)^2+G^c(s)\right]\lambda(-s)\,ds,\quad \text{for}\quad 0\le y\le t_0.
$$

Now plugging (5.1) and (C.1) into (3.23) yields that

$$\begin{aligned}\sigma^2_{\hat{X},o}(t) &= \int_0^{t_0} H_u(t)H_u^c(t)G^c(u)\lambda(-u)\,du \\ &\quad + \int_0^{t_0} (H_u^c(t))^2\left[(c_\lambda^2-1)G^c(u)^2+G^c(u)\right]\lambda(-u)\,du \\ &= \int_0^{t_0} H_u(t)G^c(t+u)\lambda(-u)\,du \\ &\quad + \int_0^{t_0} H_u^c(t)G^c(t+u)\left[(c_\lambda^2-1)G^c(u)+1\right]\lambda(-u)\,du \\ &= \int_0^{t_0} G^c(t+u)\left[(c_\lambda^2-1)G^c(t+u)+1\right]\lambda(-u)\,du \\ &= \int_{-t_0}^{0} G^c(t-s)\left[(c_\lambda^2-1)G^c(t-s)+1\right]\lambda(s)\,ds,\end{aligned}$$

where the last equality holds by a change of variable. Summing the above equation with $\sigma^2_{\hat{X},\nu}(t)$ in (3.22) yields (5.4). □

Proof of Lemma 4.2. We mimic the proof of (A.14.15) in [36]. First for a deterministic $g \in \mathbb{D}_{\mathbb{D}}$, we have, for $t \in [0,T]$, $y \in [0,y^{\uparrow}]$

$$\mathbb{E}\|g-X\|_{T,y^{\uparrow}} \geq \mathbb{E}|g(t,y)-X(t,y)| \geq |g(t,y)-\mathbb{E}[X(t,y)]|,$$

where the second inequality holds by Jensen's inequality. Hence, we have

$$\mathbb{E}\|g-X\|_{T,y^{\uparrow}} \geq \sup_{(t,y)\in[0,T]\times[0,y^{\uparrow}]} |g(t,y)-\mathbb{E}[X(t,y)]| = \|g-\mathbb{E}[X]\|_{T,y^{\uparrow}}. \tag{C.2}$$

By conditioning on X, we have

$$\begin{aligned}&\mathbb{E}\left[\phi\left(\|X-X^*\|_{T,y^{\uparrow}}\right)\right] \\ &= \mathbb{E}_X\left[\mathbb{E}_{X^*}\left[\phi\left(\|X-X^*\|_{T,y^{\uparrow}}\right)\middle| X\right]\right] \geq \mathbb{E}_X\left[\phi\left(\mathbb{E}_{X^*}\left[\|X-X^*\|_{T,y^{\uparrow}}\middle| X\right]\right)\right] \\ &\geq \mathbb{E}_X\left[\phi\left(\|X-\mathbb{E}\left[X^*\right]\|_{T,y^{\uparrow}}\right)\right] = \mathbb{E}\left[\phi\left(\|X-\mathbb{E}\left[X\right]\|_{T,y^{\uparrow}}\right)\right],\end{aligned}$$

where the first inequality holds by Jensen's inequality and the second inequality holds by (C.2). □

APPENDIX D: A CHALLENGING TEST CASE

In order to more fully substantiate the theoretical results, we consider a $G_t/GI/\infty$ IS model with non-Markov arrival process, non-exponential service-time distribution and unconventional initial conditions. We let the (artificial) initial conditions be generated by a time changed renewal process.

Corollary D.1 (Initial customers generated from a renewal process with a time change). *In the* n^{th} $G_t/GI/\infty$ *system, if the initial age process* $X_n^e(0,y) \equiv N^*(nX^*(0,y))$ *where* N^* *is a rate-1 renewal process with inter-renewal SCV* $\bar{c}_0^2$ *and the deterministic function*

$$X^*(y) \equiv \int_0^y b^*(y)dy < \infty, \tag{D.1}$$

then Assumption 2 is satisfied with $X(0,y) = X^*(y)$, $\hat{X}(0,y) = \bar{c}_0 \mathcal{B}_0(X^*(y))$, *for* $y \geq 0$, *where* $\mathcal{B}_0$ *is a standard BM. In addition, the variance of the ICP (that is (3.23)), is*

$$\begin{aligned}\sigma^2_{\hat{X},o}(t) &= \int_0^\infty H_u(t)H_u^c(t)\,dX^*(u) + \bar{c}_0^2 \int_0^\infty H_u^c(t)^2\,dX^*(u)\\ &= \int_0^\infty \left[\left(\bar{c}_0^2 - 1\right)H_u^c(t) + 1\right] H_u^c(t) b^*(u)\,du.\end{aligned}$$

Example D.1 (Simulation comparison for an example of Corollary D.1). In addition to the initial conditions specified above, we let the new input come from an $H_2^t(1,4)/LN(1,4)/\infty$ model in $[0,T]$, having a G_t arrival process $N_n(t) \equiv N^{(e)}(n\Lambda(t))$, where $N^{(e)}$ is a rate-1 equilibrium renewal process having an H_2 interrenewal distribution with balanced means and $c_\lambda^2 = 4$, while $\Lambda(t) \equiv \int_0^t \lambda(u)du$, where $\lambda(t)$ is the sinusoidal arrival rate with in (5.5) having parameters $a = c = 1$, $b = 0.6$, $\phi = 0$ and $n = 100$. This example has an LN service distribution with mean $1/\mu = 1$ and $c_s^2 = 4$. For the initial conditions, let N^* be a rate-1 Poisson process (so that $\bar{c}_0^2 = 1$) in Corollary D.1 and consider two density functions in (D.1)

$$b_1^*(u) = u\,\mathbf{1}_{(0\leq u\leq d)} + (2d-u)\,\mathbf{1}_{(d\leq u\leq 2d)} \quad \text{and} \quad b_2^*(u) = \frac{1}{3}u^2\,\mathbf{1}_{(0\leq u\leq 2d)}, \tag{D.2}$$

with $d = 1.5$, as shown in Figure 2.

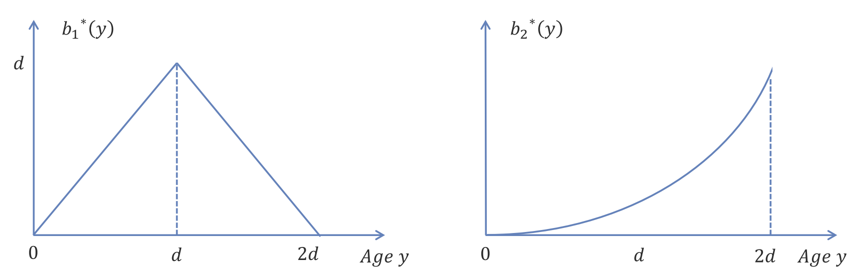

Fig 2. *Two choices of the initial-condition density* $b^*(u)$ *in* (D.1).

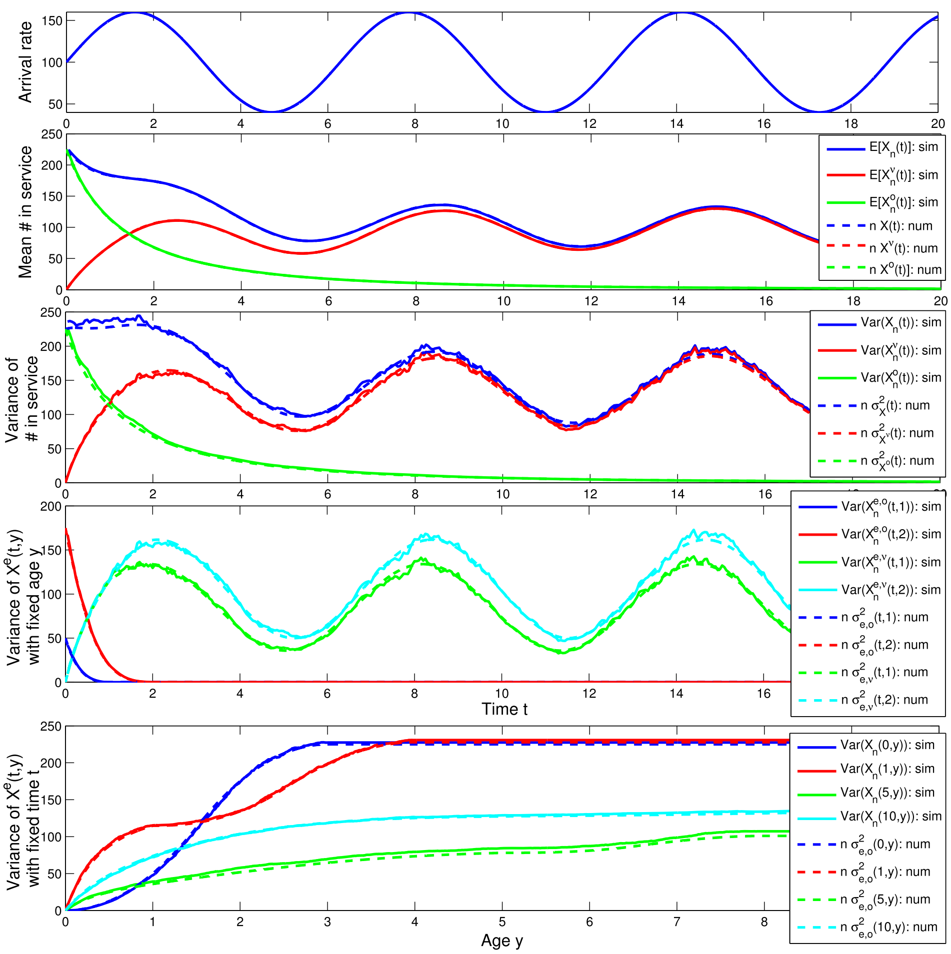

FIG 3. *Example 3 with b_1^* in* (D.2)*: Simulation comparisons of the mean and variance for the number of customers in service of an $H_2^t(1,4)/LN(1,4)/\infty$ model, with the sinusoidal arrival rate (5.5) having parameters $a = c = 1$, $b = 0.6$, $\phi = 0$ and $n = 100$ and general initial conditions.*

Comparisons with simulations are shown in Figure 3. The first three plots in Figure 3 are analogs of those in Figure 1. In the fourth and fifth plots we give simulation comparisons for the variances of the two-parameter process $\hat{X}^e(t, y)$. We provide the simulations for the case of b_2^* in the appendix.

APPENDIX E: AN ADDITIONAL EXAMPLE

EXAMPLE E.1 (Simulation comparison for another example of Corollary D.1). We now supplement Example 3 by considering that same

$H_2^t(1,4)/LN(1,4)/\infty$ model with all parameters specified in Example 3, but with $b^* = b_2^*$ specified in (D.2). Comparisons with simulations are shown in Figure 4.

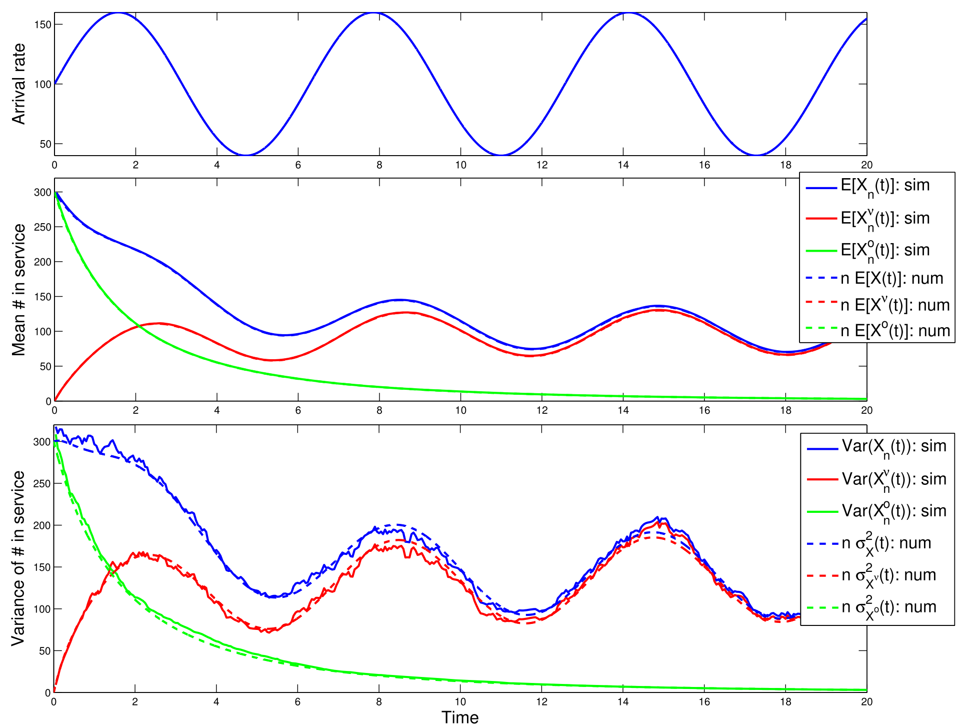

FIG 4. *Example 4 with b_2^* in (D.2): Simulation comparisons of the mean and variance for the number of customers in service of an $H_2^t(1,4)/LN(1,4)/\infty$ model, with the sinusoidal arrival rate (5.5) having parameters $a = c = 1$, $b = 0.6$, $\phi = 0$ and $n = 100$ and general initial conditions.*

APPENDIX F: STEADY-STATE APPROXIMATIONS

An important application of the results in this paper is generating useful approximations for the steady-state behavior of general stationary $G/GI/\infty$ IS models. Since we can apply Little's law to conclude that the steady-state mean number in system is $\rho \equiv \lambda E[S]$, where S is a service time, we assume that $E[S] < \infty$ in this section.

Finding general conditions for the existence of such steady-state distributions is complicated, even for the special case of the number in system with renewal (GI) arrival processes, as can be seen from Remark 2 of [40]. However, assuming that steady state for the process $\{\hat{X}_n^{e,\nu}(t,\cdot) : t \geq 0\}$ in $\mathbb{D}_{\mathbb{D}}$

for system n is well defined, it is natural to approximate by the steady-state stationary process associated with the limit process $\{\hat{X}^{e,\nu}(t,\cdot) : t \geq 0\}$ in $\mathbb{D}_{\mathbb{D}}$, which is itself an element of $\mathbb{D}$.

As observed by Glynn and Whitt [9], pp.193–195, the steady-state behavior of the limit process is relatively easy to analyze. For example, they observed, for service-time distributions with finite support in an interval $[0, y^*]$, that the number in system is in steady state after time y^*. We establish a generalization of that result, which holds because the driving processes $\mathcal{B}_a(s)$ and $\hat{K}(s,\cdot)$ have stationary increments in s.

COROLLARY F.1 (Stationary version of the limiting IS age process). *If $\hat{N}(\cdot) = c_a\sqrt{\lambda}\mathcal{B}_a(\cdot)$ and the system starts empty at time $-t_0 \leq 0$, then the limit processes (as $n \to \infty$) of the new input in the $G/GI/\infty$ model can be represented as*

$$(\hat{X}_1^{e,\nu}(t,y), \hat{X}_2^{e,\nu}(t,y)) = \left(\sqrt{\lambda c_a^2}\int_{(t-y)^+}^{t} G^c(t-s)\, d\mathcal{B}_a(s), \int_{(t-y)^+}^{t}\int_{t-s}^{\infty} d\hat{K}(\lambda s, x)\right), \tag{F.1}$$

$$(\hat{X}_1^{r,\nu}(t,y), \hat{X}_2^{r,\nu}(t,y)) = \left(\sqrt{\lambda c_a^2}\int_{-t_0}^{t} G^c(t+y-s)\, d\mathcal{B}_a(s), \int_{-t_0}^{t}\int_{t+y-s}^{\infty} d\hat{K}(\lambda s, x)\right). \tag{F.2}$$

(a) If $G^c(y^*) = 0$, then the distribution of $(\hat{X}_1^{e,\nu}(t,\cdot), \hat{X}_2^{e,\nu}(t,\cdot))$ as a process in $\mathbb{D}^2$ is independent of t for $t + t_0 \geq y^*$ and thus reaches steady state at time y^* if $t_0 = 0$ (or is in steady state at time 0 if $t_0 \geq y^*$);

(b) As $-t_0 \downarrow -\infty$, corresponding to the system starting empty in the distant past, the processes in (F.1) and (F.2) converge w.p.1 to the associated stationary processes (as functions of t)

$$(\hat{X}_1^{e,\nu}(t,y), \hat{X}_2^{e,\nu}(t,y)) = \left(\sqrt{\lambda c_a^2}\int_{t-y}^{t} G^c(t-s)\, d\mathcal{B}_a(s), \int_{t-y}^{t}\int_{t-s}^{\infty} d\hat{K}(\lambda s, x)\right), \tag{F.3}$$

$$(\hat{X}_1^{r,\nu}(t,y), \hat{X}_2^{r,\nu}(t,y)) = \left(\sqrt{\lambda c_a^2}\int_{-\infty}^{t} G^c(t+y-s)\, d\mathcal{B}_a(s), \int_{-\infty}^{t}\int_{t+y-s}^{\infty} d\hat{K}(\lambda s, x)\right), \tag{F.4}$$

whose marginal distribution as a function of y (as a process in $\mathbb{D}$) can be seen by setting $t = 0$, yielding the steady-state processes

(F.5)
$$(\hat{X}_1^{e,s}(y), \hat{X}_2^{e,s}(y)) \equiv \left(\sqrt{\lambda c_a^2}\int_{-y}^{0} G^c(-s)\, d\mathcal{B}_a(s),\ \int_{-y}^{0}\int_{-s}^{\infty} d\hat{K}(\lambda s, x)\right),$$
(F.6)
$$(\hat{X}_1^{r,s}(y), \hat{X}_2^{r,s}(y)) \equiv \left(\sqrt{\lambda c_a^2}\int_{-\infty}^{0} G^c(y-s)\, d\mathcal{B}_a(s),\ \int_{-\infty}^{0}\int_{y-s}^{\infty} d\hat{K}(\lambda s, u)\right),$$

for $y \ge 0$, with covariance formulas given by

$$\mathrm{Cov}(\hat{X}_1^{e,s}(y_1), \hat{X}_1^{e,s}(y_2)) = \lambda c_a^2 \int_0^{y_1\wedge y_2} G^c(s)^2\, ds,$$
$$\mathrm{Cov}(\hat{X}_2^{e,s}(y_1), \hat{X}_2^{e,s}(y_2)) = \lambda \int_0^{y_1\wedge y_2} G^c(s)G(s)ds,$$
$$\mathrm{Cov}(\hat{X}_1^{e,s}(y_1), \hat{X}_1^{e,s}(y_2)) = \lambda c_a^2 \int_0^{\infty} G^c(y_1+s)G^c(y_2+s)\, ds,$$
(F.7)
$$\mathrm{Cov}(\hat{X}_2^{e,s}(y_1), \hat{X}_2^{e,s}(y_2)) = \lambda \int_0^{\infty} (G^c((y_1\vee y_2)+s) - G^c(y_1+s)G^c(y_2+s))\, ds.$$

Proof. We have already established (F.1) and (F.2). The other representations hold because both $\mathcal{B}_a(s)$ and $\hat{K}(\lambda s, \cdot)$ have stationary increments in s. The limit as $-t_0 \downarrow -\infty$ is relatively easy because the processes $\mathcal{B}_a(s)$ and $\hat{K}(\lambda s, \cdot)$ do not change over the interval $[-t_0, t]$ if we expand the interval on the left and consider the process over $(-\infty, t]$. The new contribution over the interval $(-\infty, -t_0]$ decreases as $-t_0 \downarrow -\infty$. This can be quantified through the variance of the zero-mean Gaussian random variable, which is asymptotically negligible. It thus remains to derive the covariance formulas in (F.7). We only prove the first two because the proofs of the others are similar. First, by the isometry of Brownian integrals,

(F.8)
$$\mathrm{Cov}(\hat{X}_1^{e,s}(y_1), \hat{X}_1^{e,s*}(y_2)) = \lambda \int_{-y_1\wedge y_2}^{0} G^c(-s)^2\, ds = \lambda \int_0^{y_1\wedge y_2} G^c(s)^2\, ds.$$

Next, exploiting the representation of the Kiefer process in terms of the Brownian sheet, i.e., $K(x,y) = W(x,y) - yW(x,1)$ (see the appendix of [30]), we have

$$\mathrm{Cov}(\hat{X}_2^{e,s}(y_1), \hat{X}_2^{e,s}(y_2))$$
$$= \mathbb{E}\left[\int_{-y_1}^{0}\int_{-s}^{\infty} d\left(W(\lambda s, G(x)) - G(x)W(\lambda s, 1)\right)\right.$$

$$\times \int_{-y_2}^{0} \int_{-s}^{\infty} d\left(W(\lambda s, G(x)) - G(x)W(\lambda s, 1)\right)\Big]$$

(F.9)

$$= \lambda \int_{-y_1 \wedge y_2}^{0} G^c(-s)ds + \lambda \int_{-y_1 \wedge y_2}^{0} G^c(-s)^2 ds - 2\lambda \int_{-y_1 \wedge y_2}^{0} G^c(-s)^2 ds,$$

which coincides with the second covariance formula in (F.7). □

Corollary F.1 adds to the insight about the stationary model provided by Corollaries 3.1, 4.1 and 4.2 of [30]. As indicated in Corollary 4.1 of [30], we see that the approximating stationary distribution depends on the arrival process beyond its constant arrival rate λ and the assumed FCLT only through the asymptotic variability parameter c_a^2. Thus this distribution is the same as for the $M/GI/\infty$ model if and only if $c_a^2 = 1$. It is thus instructive to also consider what can be established for the $M/GI/\infty$ model directly by exploiting its special structure. So now we consider the $M/GI/\infty$ model. It is well known that the steady-state number of customers $X_n(0) \equiv X_n^e(0,\infty)$ is a Poisson random variable with $E[X_n(0)] = Var(X_n(0)) = n\lambda E[S] = n\lambda \int_0^\infty G^c(u)du$ and, conditioned on that number, the ages (and the residual service times) are i.i.d. with the stationary-excess cdf $G_e(t) \equiv (1/E[S]) \int_0^t G^c(u)\, du$, $t \geq 0$ and ccdf $G_e^c(t) \equiv 1 - G_e(t)$; e.g., see [5, 10]. Thus we have the following corollary.

Corollary F.2 (FCLT for the $M/GI/\infty$ model in steady state). *Consider a sequence of $M/GI/\infty$ models in steady state at time 0, with service-time cdf G and constant arrival rate $n\lambda$. Then Assumption 1 holds with the FWLLN and FCLT limits for the initial age processes (and all $t \geq 0$) given by*

(F.10)
$$X^{e,s}(0,y) = \rho G_e(y) = \int_0^y a(y)dy = \int_0^y \lambda G^c(u)\, du \quad \textit{and}$$
$$\hat{X}^{e,s}(0,y) \stackrel{\mathrm{d}}{=} \hat{X}_1^{e,s*}(y) + \hat{X}_2^{e,s*}(y),$$

where $\rho \equiv \lambda E[S] = X^e(0,\infty) = X(0)$, and $\hat{X}_1^{e,s*}$ and $\hat{X}_1^{e,s*}$ are independent processes with

$$\hat{X}_1^{e,s*}(y) \equiv \hat{U}^*(\rho, G_e(y)) \stackrel{\mathrm{d}}{=} \sqrt{\rho}\, \mathcal{B}_s^*(G_e(y))$$
(F.11)
$$\hat{X}_2^{e,s*}(y) \equiv G_e(y)\hat{X}(0) \stackrel{\mathrm{d}}{=} \sqrt{\rho} G_e(y)\, \mathcal{Z}_0,$$

where $\hat{U}^*$ is a standard Kiefer process associated with old customers, $\mathcal{B}_s^*$ is a standard Brownian bridge, $\mathcal{Z}_0$ is a standard Gaussian random variable, *independent of $\hat{U}^*$. The steady-state version of the remaining-service-time process*

(F.12)

$$X^{r,s}(0,y) = \rho G_e^c(y) = \int_y^\infty \lambda G^c(u)\,du \quad \textit{and} \quad \hat{X}^{r,s}(0,y) \stackrel{\mathrm{d}}{=} \hat{X}_1^{r,s*}(y) + \hat{X}_2^{r,s*}(y),$$

for $t \geq 0$, where $\hat{X}_1^{r,s*}$ and $\hat{X}_2^{r,s*}$ are independent zero-mean Gaussian processes, with

$$\mathrm{Cov}\left(\hat{X}_1^{r,s*}(x_1), \hat{X}_1^{r,s*}(x_2)\right) = \int_0^\infty H_u(x_1 \wedge x_2) H_u^c(x_1 \vee x_2)\, dX^{e,s}(0,u),$$

$$\textit{and} \quad \hat{X}_2^{r,s*}(x) \stackrel{\mathrm{d}}{=} \sqrt{\rho} \int_0^\infty H_u^c(x)\, d\mathcal{B}_o^*(G_e(u)) + \sqrt{\rho} G_e^c(y)\, \mathcal{Z}_0. \tag{F.13}$$

The variance of $\hat{X}^{e,s*}(y)$ is

$$\mathrm{Var}(\hat{X}^{e,s*}(y)) = \int_0^y \lambda G^c(u)\, du, \quad y \geq 0. \tag{F.14}$$

As a consequence, the variance of the total content, as the sum of variances of the new content (that is (3.22)) and old content (that is (3.23)), is

$$\begin{aligned} \sigma_{\hat{X}}^2 = \sigma_{\hat{X},\nu}^2 + \sigma_{\hat{X},o}^2 &= \lambda \int_0^t G^c(u)du + \lambda \int_0^\infty H_u^c(t) G^c(u)du \\ &= \lambda \int_0^\infty G^c(u)du = \rho. \end{aligned} \tag{F.15}$$

Proof. Let $A_1, A_2, \ldots$ be the ages (e.g., elapsed time in service) of the customers in service at time 0. To prove the FWLLN in (F.10), we have

$$\bar{X}^e(0,y) = \frac{1}{n} \sum_{i=1}^{n\bar{X}_n(0)} \mathbf{1}(A_i \leq y) \Rightarrow X(0)\, G_e(y) = \rho\, G_e(y) \quad \text{in } \mathbb{D}, \quad \text{as } n \to \infty,$$

where the convergence holds because the age A_i follows the equilibrium distribution G_e (see Theorem 1 of [5]). To prove the FCLT in (F.10), we have

$$\begin{aligned} \hat{X}^e(0,y) &= \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n\bar{X}_n(0)} \mathbf{1}(A_i \leq y) - nX(0)\, G_e(y) \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n\bar{X}_n(0)} (\mathbf{1}(A_i \leq y) - G_e(y)) + \hat{X}_n(0)\, G_e(y) \end{aligned} \tag{F.16}$$

$$
\begin{aligned}
&= \hat{K}_n\left(\bar{X}_n(0), G_e(y)\right) + \hat{X}_n(0)\, G_e(y) \\
&\Rightarrow \hat{U}^*\left(X(0), G_e(y)\right) + \hat{X}(0)\, G_e(y) \\
&\stackrel{\mathrm{d}}{=} \sqrt{X(0)}\mathcal{B}_s^*\left(G_e(y)\right) + \sqrt{X(0)}\, G_e(y)\, \mathcal{Z}_0 \quad \text{in } \mathbb{D},
\end{aligned}
$$

as $n \to \infty$, where the second equality holds by adding and subtracting $G_e(y)$ in the sum, and the convergence holds by (i) the marginal convergence of each of the two terms in (F.16) (due to the Gaussian CLT limit for a Poisson random variable) and by (ii) Lemma 4.1 and the conditional independence of these two terms conditioning on $X_n(0)$ (with arguments similar to §4.1). The variance formulas (F.14) and (F.15) immediately follow from (F.12) and (3.23) with $G_\nu = G$. The proof of the FWLLN limit in (F.12) follows from (F.10) and Theorem 3.1, and the proof of (F.13) follows from (F.12), Theorem 3.2 and Corollary 3.3. □

COROLLARY F.3 (Two equivalent decompositions). *For the $M/GI/\infty$ model (with $c_a^2 = 1$), the two representations (F.5) and (F.10) ((F.6) and (F.12)) are equivalent independent decompositions, i.e.,*

$$
\text{(F.17)} \quad \hat{X}_1^{e,s*} + \hat{X}_2^{e,s*} \stackrel{\mathrm{d}}{=} \hat{X}_1^{e,s} + \hat{X}_2^{e,s} \quad \text{and} \quad \hat{X}_1^{r,s*} + \hat{X}_2^{r,s*} \stackrel{\mathrm{d}}{=} \hat{X}_1^{r,s} + \hat{X}_2^{r,s}.
$$

Proof. *We only prove the first equality in (F.17) since the second equality follows similarly. Because all four terms in the first equality of (F.17) are zero-mean Gaussian processes, with $\hat{X}_1^{e,s*}$ independent of $\hat{X}_2^{e,s*}$ and $\hat{X}_1^{e,s}$ independent of $\hat{X}_2^{e,s}$, it suffices to show that*

$$
\text{(F.18)} \qquad \sum_{k=1}^{2} \mathrm{Cov}(\hat{X}_k^{e,s*}(y_1), \hat{X}_k^{e,s*}(y_2)) = \sum_{k=1}^{2} \mathrm{Cov}(\hat{X}_k^{e,s}(y_1), \hat{X}_k^{e,s}(y_2)),
$$

By (F.11), we have

$$
\begin{aligned}
&\mathrm{Cov}(\hat{X}_1^{e,s*}(y_1), \hat{X}_1^{e,s*}(y_2)) = \rho\left[G_e(y_1) \wedge G_e(y_2) - G_e(y_1)G_e(y_2)\right], \\
&\mathrm{Cov}(\hat{X}_2^{e,s*}(y_1), \hat{X}_2^{e,s*}(y_2)) = \rho G_e(y_1)G_e(y_2).
\end{aligned}
$$

so that the left-hand side of (F.18) is $\rho G_e(y_1) \wedge G_e(y_2) = \lambda \int_0^{y_1 \wedge y_2} G^c(u)\, du$, which coincides with the right-hand side of (F.18), according to (F.7) with $c_a = 1$. □

The corollaries above show that the evolution of new and old content after some time is somewhat complicated for this basic $M/GI/\infty$ model,

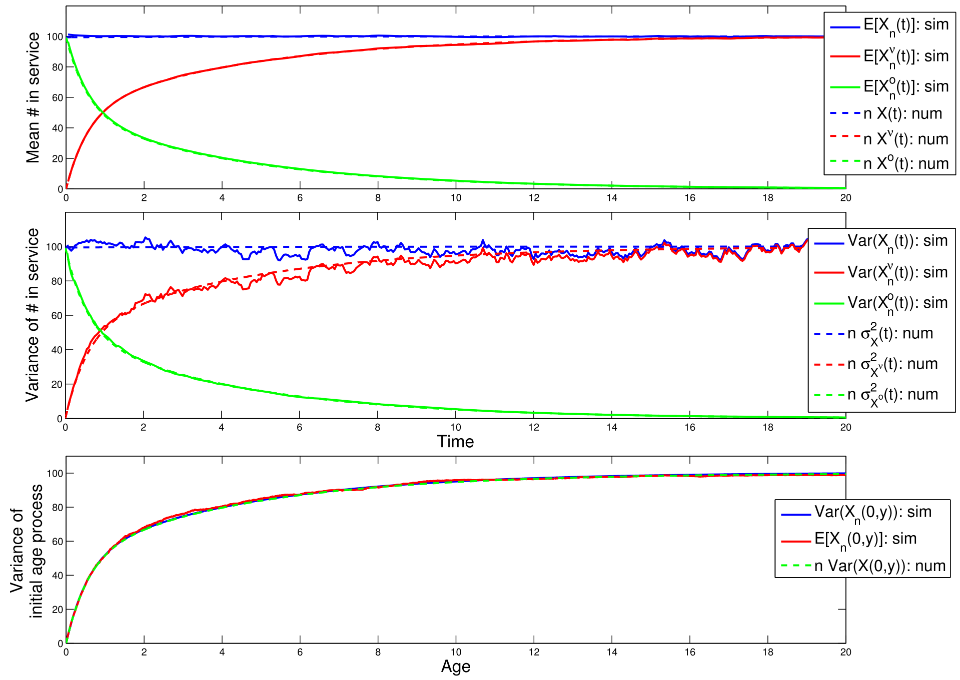

FIG 5. *Example 2: Simulation comparisons of the mean and variance for the number of new and old customers in service of an $M/H_2(1,4)/\infty$ model in steady state, with a constant arrival rate $n\lambda = 100$.*

even though the steady-state distribution at one time is remarkably simple. We now illustrate with an example.

EXAMPLE F.1 (Simulation comparison for an $M/GI/\infty$ in steady state). We consider an $M/H_2(1,4)/\infty$ model in $[0,T]$, having a Poisson arrival process with constant arrival rate $\lambda_n = n\lambda$ for $\lambda = 1$, an H_2 service distribution with balanced means,i.e., a mixture of two exponential r.v.'s with rates μ_1 and μ_2 with probability p. We set $\mu_1 = 2p\mu$, $\mu_2 = 2(1-p)\mu$, $\mu = 1$ and $p = 0.5(1-\sqrt{0.6})$ so that the mean $1/\mu = 1$ and $c_s^2 = 4$. We set $n = 100$.

Since the model is in steady state at time 0, we use a Poisson distribution with mean $n\lambda/\mu$ to generate the number of customers at time 0. To generate the elapsed times in service (e.g., ages) and the residual service times for these customers, we use the equilibrium version of the service distribution, which again follows an H_2 distribution, but with altered parameters, specifically with $p^* = p\mu_2/(p\mu_2 + (1-p)\mu_1)$, $\mu_1^* = \mu_1$ and $\mu_2^* = \mu$. See Theorem 1 of [5]. See Figure 5 for the simulation comparison, which is an analog of Figure 1.

Acknowledgments. The first two authors were supported by NSF grant CMMI 1362310 and the last author was supported by NSF grants CMMI 1066372, 1265070 and 1634133.

REFERENCES

[1] Billingsley, P(1999). *Convergence of Probability Measures*, 2nd ed., Wiley, New York. MR1700749

[2] Bremaud, P. (1981). *Point Processes and Queues: Martingale Dynamics*, Springer-Verlag, New York. MR0636252

[3] Decreusefond, L. and Moyal, P. (2008). A functional central limit theorem for the $M/GI/\infty$ queue. *Ann. Appl. Probab.* **18**(6) 2156–2178. MR2473653

[4] Donsker, M. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **23** 277–281. MR0047288

[5] Duffield, N. and Whitt, W. (1997). Control and recovery from rare congestion events in a large multi-server system. *Queueing Syst.* **26** 69–104. MR1480867

[6] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York. MR0838085

[7] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, New York. MR0270403

[8] Gerhardt, I. and Nelson, B. (2009). Transforming renewal processes for simulation of nonstationary arrival processes. *INFORMS J. Comput.* **21**(4) 630–640. MR2588345

[9] Glynn, P. W. and Whitt, W. (1991). A new view of the heavy-traffic limit theorem for the infinite-server queue. *Adv. in Appl. Probab.* **23**(1) 188–209. MR1091098

[10] Goldberg, D. and Whitt, W. (2008). The last departure time from an $M_t/G/\infty$ queue with a terminating arrival process. *Queueing Syst.* **58**(2) 77–104. MR2390269

[11] He, B., Liu, Y. and Whitt, W. (2016). Staffing a service system with non-Poisson nonstationary arrivals. *Probability in the Engineering and Information Sciences.* **30** 593–621. MR3569137

[12] Gut, A. (2005). *Probability: A Graduate Course.* Springer, New York. MR2125120

[13] Halfin, S. and Whitt, W. (1981). Heavy-traffic limits for queues with many exponential servers. *Oper. Res.* **29**(3) 567–588. MR0629195

[14] Iglehart, D. and Whitt, W. (1970). Multiple channel queues in heavy traffic II: Sequences, networks, and batches. *Adv. in Appl. Probab.* **2**(2) 355–369. MR0282443

[15] Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin. MR0959133

[16] Kang, W. and Pang, G. (2014). Equivalence of fluid models for $G_t/GI/N+GI$ queues. Working paper, Pennsylvania State University.

[17] Kang, W. and Ramanan, K. (2010). Fluid limits of many-server queues with reneging. *Ann. Appl. Probab.* **20**(6) 2204–2260. MR2759733

[18] Kaspi, H. and Ramanan, K. (2011). Law of large numbers limits for many-server queues. *Ann. Appl. Probab.* **21**(1) 33–114. MR2759196

[19] Kaspi, H. and Ramanan, K. (2013). SPDE limits of many-server queues. *Ann. Appl. Probab.* **23**(1) 145–229. MR3059233

[20] Krichagina, E. V. and Puhalskii, A. A. (1997). A heavy-traffic analysis of a closed queueing system with a GI/∞ service center. *Queueing Syst.* **25** 235–280. MR1458591

[21] Liu, R., Kuhl, M. E., Liu, Y. and Wilson, J. R. Modeling and simulation of nonstationary and non-Poisson processes. Working paper, North Carolina State University, 2017.

[22] Liu, Y. and Whitt, W. (2011). Large-time asymptotics for the $G_t/M_t/s_t + GI_t$ many-server fluid queue with abandonment. *Queueing Systems* **67** 145–182. MR2771198
[23] Liu, Y. and Whitt, W. (2012). A many-server fluid limit for the $G_t/GI/s_t + GI$ queueing model experiencing periods of overloading. *Oper. Res. Lett.* **40**(5) 307–312. MR2956434
[24] Liu, Y. and Whitt, W. (2014). Many-server heavy-traffic limit for queues with time-varying parameters. *Ann. Appl. Probab.* **24**(1) 378–421. MR3161651
[25] Liu, Y. and Whitt, W. (2014). Stabilizing performance in networks of queues with time-varying arrival rates. *Prob. Engr. Inf. Sci.* **28**(4) 419–449. MR3256197
[26] Louchard, G. (1988). Large finite population queueing systems. Part 1: The infinite server model *Comm. Statist. Stochastic Models* **4**(3) 473–505. MR0971602
[27] Marcus, M. B. and Zinn, J. (1984). The bounded law of the iterated logarithm for the weighted empirical distribution process in the non-i.i.d case. *Ann. Probab.* **12**(2) 335–360. MR0735842
[28] Massey, W. A. and Whitt, W. (1994). Unstable asymptotics for nonstationary queues. *Math. Oper. Res.* **19**(2) 267–291. MR1290501
[29] Pang, G., Talreja, R. and Whitt, W. (2007). Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probab. Surv.* **4** 193–267. MR2368951
[30] Pang, G. and Whitt, W. (2010). Two-parameter heavy-traffic limits for infinite-server queues. *Queueing Syst.* **65**(4) 325–364. MR2671058
[31] Pang, G. and Whitt, W. (2013). Two-parameter heavy-traffic limits for infinite-server queues with dependent service times. *Queueing Syst.* **73**(2) 119–146. MR3016577
[32] Puhalskii, A. A. and Reed, J. E. (2010). On many-server queues in heavy traffic. *Ann. Appl. Probab.* **20**(1) 129–195. MR2582645
[33] Puhalskii, A. A. and Reiman, M. I. (2000). The multiclass $GI/PH/N$ queue in the Halfin-Whitt regime. *Adv. in Appl. Probab.* **32**(2) 564–595. MR1778580
[34] Reed, J. (2009). The $G/GI/N$ queue in the Halfin-Whitt regime. *Ann. Appl. Probab.* **19**(6) 2211–2269. MR2588244
[35] Reed, J. and Talreja, R. (2015). Distribution-valued heavy-traffic limits for the $G/GI/\infty$ queue. *Ann. Appl. Probab.* **25**(3) 1420–1474. MR3325278
[36] Shorack, G. R. and Wellner, J. A. (2009). *Empirical Processes with Applications to Statistics.* SIAM Classics **59** (Updated version of 1986 Wiley edition). MR3396731
[37] Skorohod, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probab. Appl.* **1** 261–290. MR0084897
[38] Talreja, R. and Whitt, W. (2009). Heavy-traffic limits for waiting times in many-server queues with abandonment. *Ann. Appl. Probab.* **19**(6) 2137–2175. MR2588242
[39] Vapnik, V. N. and Cervonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications* **16** 264–280. MR0288823
[40] Whitt, W. (1982). On the heavy-traffic limit theorem for $GI/G/\infty$ queues. *Adv. Appl. Probab.* **14**(1) 171–190. MR0644013
[41] Whitt, W. (2002). *Stochastic-Process Limits*, Springer-Verlag, New York. MR1876437
[42] Whitt, W. (2005). Heavy-traffic limits for the $GI/H_2^*/n/m$ queue. *Math. Oper. Res.* **30**(1) 1–27. MR2125135
[43] Zhang, J. (2013). Fluid models of many-server queues with abandonment. *Queueing Syst.* 73(2) 147–193. MR3016578

A. Korhan Aras
Yunan Liu
Department of Industrial
and Systems Engineering
North Carolina State University
Raleigh, North Carolina 27695-7906
USA
E-mail: akaras@ncsu.edu
yliu48@ncsu.edu

Ward Whitt
Department of Industrial Engineering
and Operations Research
Columbia University
USA
E-mail: ww2040@columbia.edu