Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

### May 14, 2014

# Heavy-Traffic Limit for the Initial Content Process

A. Korhan Aras

Operations Research Program, North Carolina State University, Raleigh, NC 27695, USA, akaras@ncsu.edu

Yunan Liu

Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695-7906, USA, yliu48@ncsu.edu, http://yunanliu.wordpress.ncsu.edu

#### Ward Whitt

Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027-6699, USA, ww2040@columbia.edu, http://www.columbia.edu/~ww2040

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 heavy-traffic 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 addresses 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.

*Key words*: 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.

MSC2000 subject classification: 60F17, 60F25, 90B22

*OR/MS subject classification*: Primary: queues; secondary: limit theorems, diffusion models *History*: Submitted: May 14, 2014. Received: date. Accepted: date

1. Introduction The initial content process of a queueing system 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 initial content process is a Markov process, even if the service-time distribution is non-exponential, and thus provides a useful description of the system state at each time. In this paper we establish a heavy-traffic *functional central limit theorem* (FCLT) for the initial content process of a large-scale queueing system. The key assumption is an 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.

**Our Contribution.** We establish the FCLT above in a more general context. Specifically, we consider a sequence of infinite-server (IS) queues indexed by n, which we refer to as  $G_t/GI^o, GI^{\nu}/\infty$  models. 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, with service times distributed according to a *cumulative distribution function* (cdf) G. We assume that 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  $G_{\nu}$ . As in the usual many-server heavy-traffic scaling, the two service-time cdf's G and  $G_{\nu}$  are not scaled by n. Our approach is designed especially to treat the case in which these cdf's are *not* exponential.

We will be interested in the system performance after time 0, which can be characterized by the pair of two-parameter stochastic processes  $(X_n^{e,o}(t,y), X_n^{e,\nu}(t,y))$  with  $t \ge 0$  and  $y \ge 0$ . 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 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 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  $\nu$  new.) Given the assumptions on the service times, the stochastic process  $(X_n^{e,\nu}, X_n^{e,\nu}) \equiv \{(X_n^{e,\nu}(t,\cdot), X_n^{e,\nu}(t,\cdot)) : t \ge 0\}$  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 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 [31, 33, 39]. It is an alternative to measure-valued approaches in [5, 16, 17, 43] and distribution-valued approach in [36]. The alternative approaches are appealing for simplifying arguments and revealing structure; e.g., [36] shows that 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 [5]. 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 [15].

Theorem 2 here extends Theorems 3.2 and 5.1 of Pang and Whitt [31] by treating more general initial conditions. In particular, in §5 of [31] the remaining service times of customers initially in the system at time 0 are assumed to come from a sequence of i.i.d. random variables; similar restrictive conditions on the initial conditions are made in [17, 34, 35]. Our main contribution here is treating the *initial content process* (ICP)  $X_n^{e,o}$ ; the limit for  $X_n^{e,\nu}$  comes from [31]. As in [31], and in Louchard [26] and Krichagina and Puhalskii [19] before, we work with the empirical process of the service times. The key assumption for Theorem 2 here is that the initial age process  $\{X_n^{e,o}(0,y): y \ge 0\}$  (which is just the ICP at time 0) satisfies an FCLT jointly with the arrival process after time 0; see Assumption 1. To address the technical challenge of non-identically distributed remaining service times, we draw on Chapter 25 of Shorack and Wellner [37], which in turn uses a symmetrization argument from Marcus and Zinn [28]. 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 point out that Reed and Talreja [36] also treat the ICP as part of their new approach to obtaining heavy-traffic limits for the  $G/GI/\infty$  model, but the specific conditions assumed, as well

as the framework, are different. We emphasize engineering relevance, e.g., by providing an explicit characterization of the limit process, exposing key structure (see Remark 7) 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 the simulation experiments at the end).

**Motivation.** Very broadly, IS models are important to represent the time-varying offered load, i.e. the time-varying demand, in a service system, by going beyond the time-varying arrival rate to include the service times in a useful way; see [12, 22, 42]. Directly including the service times is important when the service times are relatively long. Then the load through time may not be well represented by the arrival rate function alone, because the customers are in the system for a substantial time after their arrival epochs. The expected number of customers in an infinite-server queue is referred to as the offered load because it represents the expected number of servers that would actually be used as a function of time if there were no limit on their availability. The analysis here may usefully supplement previous work on the offered load, because to understand and predict the demand in a service system, it may be useful to partition the demand into the demand resulting from new input after some time and the demand resulting from customers already in service at that time. The FCLT here may be useful to predict the future demand based on observation of the initial content process.

More narrowly, within the class of IS models, the present paper evidently is the first to establish a heavy-traffic limit for the performance of the  $G_t/GI^o, GI^\nu/\infty$  model with general nonstationary arrival process after time 0 when the system started some time in the past and there is a system change at time 0, possibly leading to a new arrival process and a new service-time distribution after time 0. Corollary 6 in §5 shows that Theorem 2 can be applied to establish an FCLT for the  $G_t/GI^o, GI^\nu/\infty$  model, even if the general initial conditions here hold before time 0.

Even more narrowly, for the  $G_t/GI/\infty$  model, it may be important to consider IS models with initial conditions more general than assumed in §5 of [31], where the remaining service times conditioned on the initial number in service was assumed to be i.i.d.

For finite-server queueing systems, to understand the performance, it may be useful to focus on the evolution of the content that is initially in service at some time, ignoring the new input after that time. That necessarily will be the case in service systems that provide service over normal working hours each day, if the system shuts down at some time, except that all customers already in service at the termination time are allowed to complete their service. For example, terminating systems with nonhomogeneous Poisson arrival processes were considered in [11]. As discussed in [6], such performance descriptions can be useful to develop effective measures to recover after system overloads.

Finally, it is important to note that IS models are useful for studying finite-server systems. First, results for IS models can be used to develop engineering approximations for finite-server models; e.g., see the peakedness approximations in [20, 32] and references therein. Second, heavy-traffic limits for IS models can be fruitfully applied to establish associated many-server heavy-traffic limits for finite-server systems; see [1, 23, 24, 27, 34, 35].

In fact, we intend to apply the results here in [2] to establish a FCLT for the  $G_t/GI/s_t+GI$  model with alternating overloaded (OL) and underloaded (UL) intervals, time-varying staffing as well as arrival rate and customer abandonment (the +GI), extending the FCLT for the  $G_t/M/s_t + GI$ model in [24]. We intend to apply the present results 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 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]. 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 many-server heavy-traffic 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 [31]. In §6 we provide the first characterization of the steady-state distribution of the new and old content in the stationary  $G/GI/\infty$  model. In §7 we report simulation results for a challenging test case having non-Markov arrival process, non-exponential service-time distribution, and general initial conditions. Finally, we provide concluding remarks in §8. Additional material appears in an appendix.

2. The Model After Time 0 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  $t \ge 0$ , 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  $\S1$ , it is important to treat the old and new customers separately. As in [31], for the new arrivals we have

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

$$X_n^{r,\nu}(t,y) = \sum_{i=1}^{N_n(t)} \mathbf{1}(A_i^{(n)} + S_i > t + y), \quad t \ge 0, \ y \ge 0$$
(2)

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*<sup>th</sup> customer has been in service (age in service) at time 0 in system *n*. Without loss of generality we assume the ages are ordered  $0 \le \tau_{n,1} \le \tau_{n,2} \le \ldots$ . Then

$$X_{n}^{e,o}(t,y) = \sum_{\substack{i=1\\ Y_{n}(0)}}^{X_{n}^{e}(0,(y-t)^{+})} \mathbf{1}(\eta_{i}(\tau_{n,i}) > t), \quad t \ge 0, \ y \ge 0,$$
(3)

$$X_n^{r,o}(t,y) = \sum_{i=1}^{X_n(0)} \mathbf{1}(\eta_i(\tau_{n,i}) > t+y), \quad t \ge 0, \ y \ge 0,$$
(4)

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  $\{\tau_{n,i} : i \ge 1\}$ , the sequence  $\{\eta_i(\tau_{n,i}) : i \ge 1\}$  is a sequence of mutually independent random variables with conditional tail probabilities

$$P(\eta_i(x) > t | \tau_{n,i} = x) \equiv H_x^c(t) \equiv 1 - H_x(t) \equiv \frac{G^c(t+x)}{G^c(x)},$$
(5)

for  $x \ge 0$ ,  $t \ge 0$ , 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  $\{\tau_{n,i} : i \ge 1\}$ , the random variables  $\eta_i(\tau_{n,i})$  are not identically distributed.

Given the processes (1)-(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

$$\bar{N}_n(t) \equiv N_n(t)/n, \quad \bar{X}_n^e(t,y) \equiv X_n^e(t,y)/n, \quad \bar{D}_n(t) \equiv D_n(t)/n, \quad \bar{X}_n^r(t,y) \equiv X_n^r(t,y)/n.$$
(6)

Let the associated CLT-scaled processes be

$$\hat{N}_{n}(t) \equiv \frac{N_{n}(t) - n\Lambda(t)}{\sqrt{n}}, \quad \hat{X}_{n}^{e}(t,y) \equiv \frac{X_{n}^{e}(t,y) - nX^{e}(t,y)}{\sqrt{n}}, \\
\hat{D}_{n}(t) \equiv \frac{D_{n}(t) - nD(t)}{\sqrt{n}}, \quad \hat{X}_{n}^{r}(t,y) \equiv \frac{X_{n}^{r}(t,y) - nX^{r}(t,y)}{\sqrt{n}},$$
(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 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  $d_{J_1}$  [7, 14, 38, 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 [38] original framework; see [31, 39] 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 [3, 7, 14, 41] for tightness criteria in  $\mathbb{D}$ , Theorem 6.2 of [31] for tightness criteria in  $\mathbb{D}_{\mathbb{D}}$ .

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 3 and 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 (7) jointly satisfy the FCLT

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

$$\tag{8}$$

where  $\hat{X}_n^e(0,\cdot)$  and  $\hat{N}$  are two independent zero-mean continuous Gaussian processes. We assume that the deterministic centering terms in (7), which come from the associated functional weak law of large numbers (FWLLN) stated below in (10), can be represented as

$$X^{e}(0,x) = \int_{0}^{x} a(u) \, du, \quad x \ge 0, \quad and \quad \Lambda(t) = \int_{0}^{t} \lambda(u) \, du, \quad t \ge 0, \tag{9}$$

where the fluid initial age density a(x) and arrival rate function  $\lambda(t)$  in (9) are nonnegative realvalued functions that are integrable over all bounded intervals.

REMARK 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  $\bar{X}_n^e(0, \cdot)$ , i.e.,

$$\left(\bar{X}_{n}^{e}(0,\cdot),\bar{N}_{n},\bar{X}_{n}(0)\right) \Rightarrow \left(X^{e}(0,\cdot),\Lambda,X(0)\right) \quad \text{in} \quad \mathbb{D}^{2} \times \mathbb{R} \quad \text{as} \quad n \to \infty, \tag{10}$$

REMARK 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 (14) below. Extensions are possible, as illustrated by  $\S10$  of [24].

REMARK 3. (Independence of the limiting initial content and new input) If the arrival process is a nonhomogeneous Poisson process, then we have an  $M_t/GI^o, GI^\nu/\infty$  IS model, for which 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. Nevertheless, Assumption 1 is very reasonable. 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 [31] 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 [29]. With either an ordinary or equilibrium renewal process, the limit process will be  $N(t) = c_{\lambda} \mathcal{B}_{a}(\Lambda(t))$ , where  $\mathcal{B}_{a}$ is a standard BM,  $\Lambda(t)$  is the deterministic time transformation, corresponding to the limiting cumulative arrival rate function and  $c_{\lambda}^2$  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 [29], see [9, 21] and §5.4 of [25] for uses of this representation of nonstationary non-Poisson arrival processes.

REMARK 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 6 and Lemma 4.

#### ASSUMPTION 2. (Regularity conditions for service-time cdf's)

The two service-time cdf's G and  $G_{\nu}$  are assumed to be continuous. In addition, the cdf G has a probability density function (pdf) g satisfying  $0 < g(x) \le g^{\uparrow} \equiv \sup_{x>0} g(x) < \infty$  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..

Note that Assumption 3 is quite reasonable, because it would be automatically satisfied if the system started empty at some finite time in the past.

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 [31] by treating more general initial conditions. In particular, the results for the new arrivals come from [31], but unlike §5 of [31], 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 1. (FWLLN) Consider the sequence of  $G_t/GI^o, GI^\nu/\infty$  queues satisfying all assumptions in §2. As  $n \to \infty$ ,

$$\left(\bar{N}_n, \bar{X}_n^e(0, \cdot), \bar{X}_n^r(0, \cdot), \bar{X}_n^e, \bar{X}_n^r, \bar{X}_n, \bar{D}_n\right) \Rightarrow \left(\Lambda, X^e(0, \cdot), X^r(0, \cdot), X^e, X^r, X, D\right)$$
(11)

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

$$\begin{aligned} X^{e}(t,y) &\equiv X^{e,o}(t,y) + X^{e,\nu}(t,y), \quad X^{r}(t,y) \equiv X^{r,o}(t,y) + X^{r,\nu}(t,y), \\ X^{e,o}(t,y) &= \int_{0}^{(y-t)^{+}} a(x) H_{x}^{c}(t) dx, \quad X^{e,\nu}(t,y) \equiv \int_{(t-y)^{+}}^{t} G_{\nu}^{c}(t-s)\lambda(s) ds, \\ X^{r,o}(t,y) &= \int_{0}^{\infty} a(x) H_{x}^{c}(t+y) dx, \quad X^{r,\nu}(t,y) \equiv \int_{0}^{t} G_{\nu}^{c}(t+y-s)\lambda(s) ds, \\ D(t) &= \Lambda(t) - X(t) = \int_{0}^{\infty} a(x) H_{x}(t) dx + \int_{0}^{t} G_{\nu}(t-s)\lambda(s) ds \end{aligned}$$
(12)

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  $a \lor b \equiv \max\{a, b\}$  and  $a \land b \equiv \min\{a, b\}$ .

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

$$\left(\hat{N}_n, \hat{X}_n^e(0, \cdot), \hat{X}_n^r(0, \cdot), \hat{X}_n^e, \hat{X}_n^r, \hat{X}_n, \hat{D}_n\right) \Rightarrow \left(\hat{N}, \hat{X}^e(0, \cdot), \hat{X}^r(0, \cdot), \hat{X}^e, \hat{X}^r, \hat{X}, \hat{D}\right)$$
(13)

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

$$\hat{X}^{e}(t,y) = \hat{X}_{1}^{e,\nu}(t,y) + \hat{X}_{2}^{e,\nu}(t,y) + \hat{X}_{1}^{e,o}(t,y) + \hat{X}_{2}^{e,o}(t,y),$$
(14)

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,

$$\hat{X}_{1}^{e,\nu}(t,y) \equiv \int_{(t-y)^{+}}^{t} G_{\nu}^{c}(t-s) \, d\hat{N}(s), \tag{15}$$

$$\hat{X}_{2}^{e,\nu}(t,y) \equiv \int_{(t-y)^{+}}^{t} \int_{0}^{\infty} \mathbf{1}(x > t-s) \, d\hat{K}_{\nu}(\Lambda(s),x), \tag{16}$$

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

$$C_{1}^{e,o}((t_{1},y_{1}),(t_{2},y_{2})) \equiv \operatorname{Cov}\left(\hat{X}_{1}^{e,o}(t_{1},y_{1}),\hat{X}_{1}^{e,o}(t_{2},y_{2})\right) = \int_{0}^{(y_{1}-t_{1})^{+} \wedge (y_{2}-t_{2})^{+}} H_{u}(t_{1} \wedge t_{2})H_{u}^{c}(t_{1} \vee t_{2}) \, dX^{e}(0,u)$$

$$(17)$$

and  $\hat{X}_{2}^{e,o}$  has the representation

$$\hat{X}_{2}^{e,o}(t,y) \equiv \int_{0}^{(y-t)^{+}} H_{x}^{c}(t) \, d\hat{X}^{e}(0,x) \equiv H_{(y-t)^{+}}^{c}(t) \hat{X}^{e}(0,(y-t)^{+}) - \int_{0}^{(y-t)^{+}} \hat{X}^{e}(0,u-) dH_{u}^{c}(t), \quad (18)$$

where  $(X^e(0, \cdot), \hat{X}^e(0, \cdot))$  is the limit of the initial age process in Assumption 1 and (10). The joint limit (13) follows from the displayed limit. The other limit processes  $\hat{X}$ ,  $\hat{D}$  and  $\hat{X}^r$  are specified in the corollaries below.

REMARK 5. (Correction in [31]) The limits for the new input follow from [31], so the formulas in (15) and (16) should be consistent with [31]. 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 [31] all should be  $\infty$  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 [31] also should be  $\infty$  instead of t. After this correction, the formulas in (16) and elsewhere are consistent with [31].

We next characterize all the other limit processes using the limit in (14). Let  $\stackrel{d}{=}_{t}$  denote equal in distribution for each t. Let  $\mathcal{B}_{s}(\cdot)$  be an independent BM (associated with service times of new customers).

COROLLARY 1. (Limits for the one-parameter queue length process) Under the assumptions of Theorem 2, the limit for the total number in service at t is

$$\hat{X}(t) \equiv \hat{X}^{e}(t,\infty) \equiv \hat{X}_{1}^{\nu}(t) + \hat{X}_{2}^{\nu}(t) + \hat{X}_{1}^{o}(t) + \hat{X}_{2}^{o}(t),$$
(19)

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

$$\hat{X}_{1}^{\nu}(t) \equiv \hat{X}_{1}^{e,\nu}(t,\infty) \equiv \int_{0}^{t} G_{\nu}^{c}(t-s) \, d\hat{N}(s), \qquad (20)$$

$$\hat{X}_{2}^{\nu}(t) \equiv \hat{X}_{2}^{e,\nu}(t,\infty) \equiv \int_{0}^{t} \int_{0}^{\infty} \mathbf{1}(x > t - s) \, d\hat{K}_{\nu}(\Lambda(s), x) \stackrel{\mathrm{d}}{=}_{t} - \int_{0}^{t} \sqrt{G_{\nu}(t - s)G_{\nu}^{c}(t - s)} d\mathcal{B}_{s}(\Lambda(s)), \tag{21}$$

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

$$C_1^o(t,t') \equiv \operatorname{Cov}\left(\hat{X}_1^o(t), \hat{X}_1^o(t')\right) = C_1^{e,o}((t_1,\infty), (t_2,\infty)) = \int_0^\infty H_u(t \wedge t') H_u^c(t \vee t') \, dX^e(0,u), \quad (22)$$

and

$$\hat{X}_{2}^{o}(t) \equiv \hat{X}_{2}^{e,o}(t,\infty) \equiv \int_{0}^{\infty} H_{x}^{c}(t) \, d\hat{X}^{e}(0,x).$$
(23)

COROLLARY 2. (Limits for the one-parameter departure process) Under the assumptions of Theorem 2, The limit for the number of departures by t is

$$\hat{D}(t) = \hat{D}_1^{\nu}(t) + \hat{D}_2^{\nu}(t) + \hat{D}_1^{o}(t) + \hat{D}_2^{o}(t), \qquad (24)$$

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

$$\hat{D}_{1}^{\nu}(t) \equiv \int_{0}^{t} G_{\nu}(t-s) \, d\hat{N}(s), \tag{25}$$

$$\hat{D}_{2}^{\nu}(t) \equiv \int_{0}^{t} \int_{0}^{\infty} \mathbf{1}(x \le t - s) \, d\hat{K}_{\nu}(\Lambda(s), x) \stackrel{\mathrm{d}}{=}_{t} \int_{0}^{t} \sqrt{G_{\nu}(t - s)G_{\nu}^{c}(t - s)} d\mathcal{B}_{s}(\Lambda(s)), \tag{26}$$

 $\hat{D}_1^o(t) = -\hat{X}_1^o(t)$  being a zero-mean Gaussian process with covariance function

$$\operatorname{Cov}\left(\hat{D}_{1}^{o}(t),\hat{D}_{1}^{o}(t')\right) = C_{1}^{o}(t,t') \quad and \quad \hat{D}_{2}^{o}(t) \equiv \hat{X}(0) - \hat{X}_{2}^{o}(t) = \int_{0}^{\infty} H_{x}(t) \, d\hat{X}^{e}(0,x). \tag{27}$$

COROLLARY 3. (Limits for the remaining-service-time process) Under the assumptions of Theorem 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  $x \ge 0$  and

$$\hat{X}^{r}(t,x) = \hat{X}_{1}^{r,\nu}(t,x) + \hat{X}_{2}^{r,\nu}(t,x) + \hat{X}_{1}^{r,o}(t,x) + \hat{X}_{2}^{r,o}(t,x),$$
(28)

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

$$\hat{X}_{1}^{r,\nu}(t,x) \equiv \int_{0}^{t} G_{\nu}^{c}(t+x-s) \, d\hat{N}(s), \tag{29}$$

$$\hat{X}_{2}^{r,\nu}(t,x) \equiv \int_{0}^{t} \int_{0}^{\infty} \mathbf{1}(u+s>t+x) \, d\hat{K}_{\nu}(\Lambda(s),u), \tag{30}$$

$$\hat{X}_{1}^{r,o}(t,x) \equiv \hat{X}_{1}^{o}(t+x) = \hat{X}_{1}^{e,o}(t+x,\infty) \quad and \quad \hat{X}_{2}^{r,o}(t,x) \equiv \hat{X}_{2}^{o}(t+x) = \hat{X}_{2}^{e,o}(t+x,\infty).$$
(31)

REMARK 6. (The stochastic integrals) The integrals in Theorem 2 should be interpreted just as in [31], as explained in Remark 3.2 there. In particular, the deterministic integrals in (17) and (22) are all Stieltjes integrals, while the integrals in (16), (21), (26) and (30) are all stochastic integrals, just as in [31]. As in Theorem 3.2 and Remark 3.3 of [31], the continuity assumption on the cdf  $G_{\nu}$  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 (15), (20), (25) and (29), and with respect to  $\hat{X}^e(0, \cdot)$  in (18), (23) and (27). As explained in Remark 3.2 of [31], these all should be interpreted as the form after the representation of integration by parts, as given on p. 336 of [4]. That is justified because the pre-limit processes of the integrator process have sample paths of bounded variation. For example, the alternative representation for (18) is given there; see §4.3.

REMARK 7. (Four independent stochastic effects) The expression for the limit process  $\hat{X}^e$  in (14) 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 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.

If we make an additional assumption for N, we can exhibit covariance and variance formulas.

COROLLARY 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

$$C^{e}((t_{1}, y_{1}), (t_{2}, y_{2})) \equiv \operatorname{Cov}(\hat{X}^{e}(t_{1}, y_{1}), \hat{X}^{e}(t_{2}, y_{2})) = C^{e,\nu}((t_{1}, y_{1}), (t_{2}, y_{2})) + C^{e,o}((t_{1}, y_{1}), (t_{2}, y_{2})),$$

where

$$C^{e,\nu}((t_1, y_1), (t_2, y_2)) \equiv \int_{(t_1 - y_1)^+ \wedge (y_2 - y_2)^+}^{t_1 \wedge t_2} \left[ (c_a^2 - 1) G_{\nu}^c(t_1 - s) G_{\nu}^c(t_2 - s) + G_{\nu}^c((t_1 \vee t_2) - s) \right] \lambda(s) ds,$$
  

$$C^{e,o}((t_1, y_1), (t_2, y_2)) \equiv \int_{0}^{(y_1 - t_1)^+ \wedge (y_2 - t_2)^+} H_u(t_1 \wedge t_2) H_u^c(t_1 \vee t_2) dX^e(0, u)$$
  

$$+ \int_{0}^{(y_1 - t_1)^+ \wedge (y_2 - t_2)^+} H_u^c(t_1) H_u^c(t_2) d\Sigma_2^{e,o}(u).$$

so that the variances are

$$\begin{split} \sigma_{e}^{2}(t,y) &\equiv \operatorname{Var}(\hat{X}^{e}(t,y)) = \sigma_{e,\nu}^{2}(t,y) + \sigma_{e,o}^{2}(t,y), \\ where \quad \sigma_{e,\nu}^{2}(t,y) = \sigma_{\nu}^{2}(((t-y)^{+},t), \\ \sigma_{\nu}^{2}(u,v) &\equiv \int_{u}^{v} \left[ (c_{a}^{2}-1)G_{\nu}^{c}(v-s)^{2} + G_{\nu}^{c}(v-s) \right] \lambda(s) ds \\ and \quad \sigma_{e,o}^{2}(t,y) &= \int_{0}^{(y-t)^{+}} H_{u}(t)H_{u}^{c}(t) dX_{0}^{e}(u) + \int_{0}^{(y-t)^{+}} H_{x}^{c}(t)^{2} d\Sigma_{2}^{e,o}(x). \end{split}$$

COROLLARY 5. (Variance for  $\hat{X}(t)$  and  $\hat{D}(t)$ ) Under the assumptions of Corollary 4, the variances of the one-parameter processes  $\hat{X}(t)$  and  $\hat{D}(t)$  are

$$\sigma_{\hat{X}}^{2}(t) \equiv \operatorname{Var}(\hat{X}(t)) = \sigma_{\hat{X},\nu}^{2}(t) + \sigma_{\hat{X},o}^{2}(t), \tag{32}$$

where 
$$\sigma_{\hat{X},\nu}^{2}(t) \equiv \sigma_{e,\nu}^{2}(t,\infty) = \sigma_{\nu}^{2}(0,t) = \int_{0}^{\infty} \left[ (c_{a}^{2}-1)G_{\nu}^{c}(t-s)^{2} + G_{\nu}^{c}(t-s) \right] \lambda(s) ds$$
 (33)

and 
$$\sigma_{\hat{X},o}^{2}(t) = \sigma_{e,o}^{2}(t,\infty) = \int_{0}^{\infty} H_{u}(t)H_{u}^{c}(t) dX_{0}^{e}(u) + \int_{0}^{\infty} H_{x}^{c}(t)^{2} d\Sigma_{2}^{e,o}(x),$$
 (34)  
 $t) = V_{0}r(\hat{D}(t)) = \sigma^{2}(t) + \sigma^{2}(t)$  (25)

$$\sigma_{\hat{D}}^{2}(t) \equiv \operatorname{Var}(\hat{D}(t)) = \sigma_{\hat{D},\nu}^{2}(t) + \sigma_{\hat{D},o}^{2}(t),$$
(35)

where 
$$\sigma_{\hat{D},\nu}^2(t) = \int_0^{\infty} \left[ (c_a^2 - 1)G_{\nu}^2(t-s) + G_{\nu}(t-s) \right] \lambda(s) ds$$
  
and  $\sigma_{\hat{D},o}^2(t) = \int_0^{\infty} H_u(t)H_u^c(t) dX_0^e(u) + \int_0^{\infty} H_x(t)^2 d\Sigma_2^{e,o}(x) dx_0^{e,o}(t) dx_0^{e,o}(t) + \int_0^{\infty} H_x(t)^2 d\Sigma_2^{e,o}(t) dx_0^{e,o}(t) dx_0$ 

REMARK 8. (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 [31]). 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 2. We start with the FCLT for all processes related to new arrivals from [31], obtaining

$$(\hat{N}_{n}, \bar{N}_{n}, \hat{K}_{n}, \bar{K}_{n}, \hat{R}_{n}, \hat{X}_{n}^{e,\nu}, \hat{X}_{n}^{r,\nu}, \hat{D}_{n}^{\nu}) \Rightarrow (\hat{N}, \bar{N}, \hat{K}, \bar{K}, \hat{R}, \hat{X}^{e,\nu}, \hat{X}^{r,\nu}, \hat{D}^{\nu}),$$
(36)

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

$$(\hat{X}_{n}(0,\cdot),\bar{X}_{n}(0,\cdot),\hat{X}_{n}^{e,o},\hat{X}_{n}^{r,o},\hat{D}_{n}^{o}) \Rightarrow (\hat{X}(0,\cdot),\bar{X}(0,\cdot),\hat{X}^{e,o},\hat{X}^{r,o},\hat{D}^{o}).$$
(37)

in  $\mathbb{D}^3 \times \mathbb{D}^2_{\mathbb{D}}$ . We will then have the joint convergence of (36) and (37) 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  $\hat{X}_n^{e,o}$  can be decomposed into two other two-parameter processes  $\hat{X}_{n,1}^{e,o}(t,y)$  and  $\hat{X}_{n,2}^{e,o}(t,y)$ , 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**  $\hat{X}_{n}^{e,o}$ . To prove (37), we use a convenient representation of  $\hat{X}_{n}^{e,o}(t,y)$ . Let  $\hat{\eta}_{n,i}(t) \equiv \mathbf{1}\{\eta_i(\tau_{n,i}) > t\} - H^c_{\tau_{n,i}}(t)$ . From (3), (7) and (12), we can write

$$\hat{X}_{n}^{e,o}(t,y) = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{X_{n}^{e}(0,(y-t)^{+})} \mathbf{1}\{\eta_{i}(\tau_{n,i}) > t\} - \int_{0}^{(y-t)^{+}} b(0,x) H_{x}^{c}(t) dx \right) \\
= \frac{1}{\sqrt{n}} \sum_{i=1}^{X_{n}^{e}(0,(y-t)^{+})} \left( \mathbf{1}\{\eta_{i}(\tau_{n,i}) > t\} - H_{\tau_{ni}}^{c}(t) \right) \\
+ \sqrt{n} \left( \int_{0}^{(y-t)^{+}} H_{x}^{c}(t) d\bar{X}_{n}^{e}(0,x) - \int_{0}^{(y-t)^{+}} b(0,x) H_{x}^{c}(t) dx \right) \\
= \frac{1}{\sqrt{n}} \sum_{i=1}^{X_{n}^{e}(0,(y-t)^{+})} \widehat{\eta}_{n,i}(t) + \int_{0}^{(y-t)^{+}} H_{x}^{c}(t) d\hat{X}_{n}^{e}(0,x) \equiv \hat{X}_{n,1}^{e,o}(t,y) + \hat{X}_{n,2}^{e,o}(t,y), \quad (38)$$

where the second equality holds by adding and subtracting  $H_{\tau_{n,i}}(t)$  in the sum. Even though  $\hat{X}_{n,1}^{e,o}$ and  $\hat{X}_{n,2}^{e,o}$  in (38) are not independent, because they both involve the age sequence  $\{\tau_{n,j}: j \geq 1\}$ or equivalently the counting process  $X_n^e(0,\cdot)$ , they are conditionally independent given  $\bar{X}_n^e(0,\cdot)$ . 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 [30]. The spaces are different here, but the argument is the same.

LEMMA 1. Let  $\{Y_n : n \ge 1\}$  and Y be processes with sample paths in  $\mathbb{D}_{\mathbb{D}}$ , and let  $\{Z_n : n \ge 1\}$ 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

$$Y_n^{Z_n} \Rightarrow Y^Z \quad in \quad \mathbb{D}_{\mathbb{D}} \quad whenever \quad Z_n \to Z \quad in \quad \mathbb{D} \quad as \quad n \to \infty \quad w.p.1, \tag{39}$$

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

We apply Lemma 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

$$\widetilde{X}_{n}^{e}(0,\cdot) \to \widetilde{X}^{e}(0,\cdot) \quad \text{in} \quad \mathbb{D} \quad \text{w.p.1.}$$
(40)

It remains to establish the limit (39), assuming (40).

Given that we condition with respect to the ages and then uncondition, in order to establish the joint convergence  $(\hat{X}_{n,1}^{e,o}, \hat{X}_{n,2}^{e,o}, \hat{X}_n^e(0, \cdot), \bar{X}_n^e(0, \cdot)) \Rightarrow (\hat{X}_1^{e,o}, \hat{X}_2^{e,o}, \hat{X}_0^e, X_0^e)$  in  $\mathbb{D}_{\mathbb{D}}^2 \times \mathbb{D}^2$ , it suffices to prove  $(\hat{X}_{n,1}^{e,o}, \bar{X}_n^e(0, \cdot)) \Rightarrow (\hat{X}_1^{e,o}, X_0^e)$  in  $\mathbb{D}_{\mathbb{D}} \times \mathbb{D}$  and  $(\hat{X}_{n,2}^{e,o}, \hat{X}_n^e(0, \cdot), \bar{X}_n^e(0, \cdot)) \Rightarrow (\hat{X}_2^{e,o}, \hat{X}_0^e, X_0^e)$ 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 (38).

4.2. Convergence of the First Term in (38). In addition to the conditioning discussed above, we use the compactness approach to prove (39) in order to establish convergence of the first term in (38); 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 (17) in Theorem 2. In Step 2 (§4.2.2), using the convergence of the covariance functions, we establish the convergence of the fidis of  $\hat{X}_{n,1}^{e,o}$  in  $\mathbb{D}_{\mathbb{D}}$ , which is equivalent to the joint convergence of  $\left(\hat{X}_{n,1}^{e,o}(t_1,\cdot),\ldots,\hat{X}_{n,1}^{e,o}(t_k,\cdot)\right)$ 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  $\left(\hat{X}_{n,1}^{e,o}(t_1,\cdot),\ldots,\hat{X}_{n,1}^{e,o}(t_k,\cdot)\right)$  in the second argument, namely, the joint convergence of the bigger vector  $\left(\hat{X}_{n,1}^{e,o}(t_1,y_j), 1 \leq i \leq k, 1 \leq j \leq m\right)$  in  $\mathbb{R}^{k \times m}$ , for all  $m \geq 1$  and  $0 < y_1 < \cdots < y_m$ . Second, we establish the tightness of  $\left(\hat{X}_{n,1}^{e,o}(t_1,\cdot),\ldots,\hat{X}_{n,1}^{e,o}(t_k,\cdot)\right)$  in  $\mathbb{D}^k$ . In Step 3 (§4.2.3) we prove that  $\hat{X}_{n,1}^{e,o}$  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  $\bar{X}_n^e(0,\cdot)$ . Upon conditioning, the first term in (38) is a non-random sum of the independent mean-zero random variables  $\hat{\eta}_{n,i}(t)$  defined at the beginning of §4.1. Hence,

$$\mathbb{E}^{\tau} \left[ \hat{X}_{n,1}^{e,o}(t,y) \hat{X}_{n,1}^{e,o}(t',y') \right] = \frac{1}{n} \sum_{i=1}^{X_n^e(0,(y-t)^+ \wedge (y'-t')^+)} \mathbb{E}^{\tau} \left[ \widehat{\eta}_{n,i}(t) \widehat{\eta}_{n,i}(t') \right]$$
$$= \frac{1}{n} \sum_{i=1}^{X_n^e(0,(y-t)^+ \wedge (y'-t')^+)} H_{\tau_{n,i}}(t) H_{\tau_{n,i}}^c(t') = \int_0^{(y-t)^+ \wedge (y'-t')^+} H_u(t) H_u^c(t') \ d\bar{X}_n^e(0,u). \tag{41}$$

Assuming (40), which corresponds to convergence of finite measures, from (41) we have

$$\mathbb{E}^{\tau} \left[ \hat{X}_{n,1}^{e,o}(t,y) \hat{X}_{n,1}^{e,o}(t',y') \right] \to \int_{0}^{(y-t)^{+} \wedge (y'-t')^{+}} H_{u}(t) H_{u}^{c}(t') \ dX_{0}(u) \equiv K(t_{i},y_{j},t_{i'},y_{j'}).$$
(42)

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 (42) 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 (42), yielding convergence of the covariance functions after unconditioning, i.e.,

$$K_{n}(t, y, t', y') \equiv \mathbb{E}\left[\mathbb{E}^{\tau}\left[\hat{X}_{n,1}^{e,o}(t, y)\hat{X}_{n,1}^{e,o}(t', y')\right]\right] = \mathbb{E}\left[\int_{0}^{(y-t)^{+}\wedge(y'-t')^{+}} H_{u}(t)H_{u}^{c}(t') \ d\bar{X}_{n}^{e}(0, u)\right]$$
  
$$\rightarrow \int_{0}^{(y-t)^{+}\wedge(y'-t')^{+}} H_{u}(t)H_{u}^{c}(t') \ dX_{0}(u) \equiv K(t, y, t', y').$$
(43)

As an immediate consequence of (43), we have an expression for the variance functions and their convergence,

$$\sigma_n^2(t,y) \equiv K_n(t,y,t,y) = \mathbb{E}\left[\int_0^{(y-t)^+} H_u(t)H_u^c(t) \ d\bar{X}_n^e(0,u)\right] \to \int_0^{(y-t)^+} H_u(t)H_u^c(t) \ dX_0(u) \equiv \sigma^2(t,y).$$
(44)

**4.2.2.** Step 2: Convergence of the fidis in  $\mathbb{D}_{\mathbb{D}}$ . We again apply Lemma 1 and start assuming (40). Hence, for each n, we condition upon the ages.

Step 2a: Joint convergence in  $\mathbb{R}^{k \times m}$ . Fix  $m \ge 1$  and  $0 < y_1 < \cdots < y_m$ . The convergence of the fidis of the vector  $(\hat{X}_{n,1}^{e,o}(t_1, \cdot), \dots, \hat{X}_{n,1}^{e,o}(t_k, \cdot))$  in the second argument is equivalent to the joint convergence of the bigger vector  $(\hat{X}_{n,1}^{e,o}(t_i, y_j), 1 \le i \le k, 1 \le j \le m)$  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  $\{a_{i,j}\} \in \mathbb{R}, i = 1, ..., k$ and j = 1, ..., m, as  $n \to \infty$ ,

$$\sum_{i=1}^{k} \sum_{j=1}^{m} a_{i,j} \hat{X}_{n,1}^{e,o}(t_i, y_j) \Rightarrow \sum_{i=1}^{k} \sum_{j=1}^{m} a_{i,j} \hat{X}_1^{e,o}(t_i, y_j) \stackrel{\mathrm{d}}{=} \mathcal{N}(0, \Sigma) \quad \text{in} \quad \mathbb{R},$$
(45)

where the variance of the limit is

$$\Sigma \equiv \sum_{i=1}^{k} \sum_{j=1}^{m} \sum_{i'=1}^{k} \sum_{j'=1}^{m} a_{i,j} a_{i',j'} K(t_i, y_j, t_{i'}, y_{j'}).$$
(46)

To establish (45), we define the random variables

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

Since  $\tilde{Y}_{n,j}$ ,  $j \ge 1$ , are independent random variables, conditioned on  $X_n^e(0, \cdot)$ , we can rewrite the left-hand side of (45) as

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

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 1, we are assuming that (40) 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 CLT for a double sequence (triangular array) of non-identically distributed independent random variables, e.g., Theorem 7.2.4 of [13]. The variance of  $\tilde{S}_n$  is

$$\tilde{s}_{n}^{2} \equiv \operatorname{Var}(\tilde{S}_{n})) = \sum_{i=1}^{k} \sum_{j=1}^{m} \sum_{i'=1}^{k} \sum_{j'=1}^{m} a_{i,j} a_{i',j'} K_{n}(t_{i}, y_{j}, t_{i'}, y_{j'}) \rightarrow \sum_{i=1}^{k} \sum_{j=1}^{m} \sum_{i'=1}^{k} \sum_{j'=1}^{m} a_{i,j} a_{i',j'} K(t_{i}, y_{j}, t_{i'}, y_{j'}) \equiv \Sigma,$$
(47)

as  $n \to \infty$  where  $\Sigma$  is defined in (46) and the convergence follows from (43). It remains to verify the Lindeberg conditions (see (2.1)-(2.2) on p.330 of [13]) or the Lyapounov condition (see (2.20) on p.339 of [13]). 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  $\Sigma$  in (47) as  $n \to \infty$ , the Lindeberg condition is satisfied. Therefore, by Theorem 7.2.4 of [13], if (40) holds, then

$$\tilde{S}_n/\tilde{s}_n \Rightarrow \mathcal{N}(0,1) \quad \text{as } n \to \infty,$$
(48)

which together with (47), imply the desired convergence in (45) under the condition (40). Lemma 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 (40). This tightness is equivalent to the tightness of each component  $\hat{X}_{n,1}^{e,o}(t_i, \cdot)$ in  $\mathbb{D}$ , for all  $1 \leq i \leq k$ , by Theorem 11.6.7. of [41].

We prove tightness of the components by proving a stronger result. Paralleling  $\S7$  of [23], we show that

$$\hat{X}_{n,1}^{e,o}(t,y) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n\bar{X}_n^e(0,y-t)} \widehat{\eta}_{n,i}(t) \Rightarrow \mathcal{B}(\sigma_t^2(y)) \quad \text{in} \quad \mathbb{D} \quad \text{as} \quad n \to \infty \quad \text{for each fixed } t, \tag{49}$$

where  $\mathcal{B}$  is a standard BM. To prove this, we observe that, under the conditioning on the ages, the process  $\{X_{n,1}^{e,o}(t,y): y \ge 0\}$  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  $n \geq 1$ . Then we can apply the martingale FCLT in Theorem 7.1.4 of [7], exploiting the fact that the summands are independent [-1, 1]-valued zero-mean random variables. The first variance function in (44) with t fixed is the quadratic variation of  $\{\hat{X}_{n,1}^{e,o}(t,y): y \ge 0\}$  and converges to the second variance formula in (44). Hence the function  $\sigma_t^2(y)$  in (49) 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  $\{\hat{X}_{n,1}^{e,o}\}$  in  $\mathbb{D}_{\mathbb{D}}$ . To complete the proof of the convergence of the first term of (38) under condition (40), we next show that the sequence  $\{\hat{X}_{n,1}^{e,o}\}$  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 [31]. The specific conditions we establish are (50) and (61) 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 (40), that for all  $\epsilon > 0$ , there exists c > 0 such that

$$\mathbb{P}^{\tau}(\|\hat{X}_{n,1}^{e,o}\|_{T,y^{\uparrow}} > c) \le \epsilon \quad \text{for all} \quad n \ge 1,$$
(50)

where  $\|\hat{X}_{n,1}^{e,o}\|_{T,y^{\uparrow}} = \sup_{(t,y)\in[0,T]\times[0,y^{\uparrow}]} |\hat{X}_{n,1}^{e,o}(t,y)|.$ To bound the probability in (50), we apply Chernoff's inequality (e.g., see Lemma 3.1.1. of [13]), obtaining, for r > 0,

$$\mathbb{P}^{\tau}(\|\hat{X}_{n,1}^{e,o}\|_{T,y^{\uparrow}} > c) \le e^{-rc} \mathbb{E}^{\tau} \left[ e^{r \|\hat{X}_{n,1}^{e,o}\|_{T,y^{\uparrow}}} \right],$$
(51)

To bound the right side of (51), we follow the symmetrization argument used in the proof of inequality 3 on p.820 of [37] (which in turn follows Lemma 1.1 of [28]). 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  $\xi_i$  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  $\hat{X}_{n,1}^{*,e,o}$  be  $\hat{X}_{n,1}^{e,o}$  with  $\{\eta_i(\tau_{n,i})\}$ replaced by  $\{\eta_i^*(\tau_{n,i})\}$ . Let  $\mathbb{E}_{\eta}^{\tau}\mathbb{E}_{\eta}^{\tau}\mathbb{E}_{\xi}^{\tau}$  denote the expectation with respect to  $\{\xi_i\}$  conditioned on  $\{\tau_{n,i}\}, \{\eta_i\}$  and  $\{\eta_i^*\}$ ; let  $\mathbb{E}_{\eta}^{\tau}\mathbb{E}_{\eta^*}^{\tau}$  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 2. Let X and  $X^*$  be two i.i.d. elements in space  $\mathbb{D}_{\mathbb{D}}$ . Suppose a function  $\phi$  is convex and nondecreasing with domain  $[0,\infty)$ . Then

$$\mathbb{E}\left[\phi(\|X - \mathbb{E}[X]\|_{T,y^{\dagger}})\right] \le \mathbb{E}\left[\phi(\|X - X^*\|_{T,y^{\dagger}})\right].$$
(52)

We omit the proof because it is similar to that of Lemma (A.14.15) in [37], but we do give the proof in the appendix.

LEMMA 3. Let  $\zeta_i$  and  $w_i$  be real-valued numbers with  $w_i \ge 0$  for  $1 \le i \le N$  and let  $\{w_{(i)}, 1 \le i \le N\}$ 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,

$$\left|\sum_{i=1}^{N} \zeta_i \mathbf{1}(w_{(N-i+1)} > t)\right| \le \max_{1 \le j \le N} \left|\sum_{i=1}^{j} \zeta_i\right|, \quad 0 \le t \le T.$$
(53)

*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

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

We now continue to bound the right side of (51). For that purpose, define

$$\Gamma_{n,i}(t,y) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{X_n^e(0,(y-t)^+)} \xi_i \cdot \mathbf{1}(\eta_i(\tau_{n,i}) > t) \quad \text{and} \quad \Gamma_{n,i}^*(t,y) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{X_n^e(0,(y-t)^+)} \xi_i \cdot \mathbf{1}(\eta_i^*(\tau_{n,i}) > t).$$
(54)

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*<sup>th</sup> largest one.

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

$$\mathbb{E}^{\tau} \left[ \exp\left(r \|\hat{X}_{n,1}^{e,o}\|_{T,y^{\dagger}}\right) \right] \leq \mathbb{E}^{\tau} \left[ \exp\left(r \|\hat{X}_{n,1}^{e,o} - \hat{X}_{n,1}^{*,e,o}\|_{T,y^{\dagger}}\right) \right] \\
= \mathbb{E}^{\tau} \left[ \exp\left(r \|n^{-1/2} \sum_{i=1}^{N_{n}^{*}(y,t)} \left(\mathbf{1}(\eta_{i}(\tau_{n,i}) > t) - \mathbf{1}(\eta_{i}^{*}(\tau_{n,i}) > t))\right) \right\|_{T,y^{\dagger}} \right) \right] \\
= \mathbb{E}^{\tau} \left[ \exp\left(r \|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^{\dagger}} \right) \right] \\
\leq \mathbb{E}^{\tau} \left[ \exp\left(r \|\Gamma_{n,i}(t,y)\|_{T,y^{\dagger}} + r \|\Gamma_{n,i}^{*}(t,y)\|_{T,y^{\dagger}} \right) \right] \\
= \mathbb{E}_{\eta}^{\tau} \mathbb{E}_{\eta}^{\pi} \mathbb{E}_{\xi}^{\tau} \left[ \exp\left(r \|\Gamma_{n,i}(t,y)\|_{T,y^{\dagger}} + r \|\Gamma_{n,i}^{*}(t,y)\|_{T,y^{\dagger}} \right) \right] \\
\leq \mathbb{E}_{\eta}^{\tau} \left[ \mathbb{E}_{\xi}^{\tau} \exp\left(2r \|\Gamma_{n,i}(t,y)\|_{T,y^{\dagger}} \right) \right]^{1/2} \cdot \mathbb{E}_{\eta^{*}}^{\tau} \left[ \mathbb{E}_{\xi}^{\tau} \exp\left(2r \|\Gamma_{n,i}^{*}(t,y)\|_{T,y^{\dagger}} \right) \right]^{1/2} \\
= \mathbb{E}_{\eta}^{\tau} \mathbb{E}_{\xi}^{\tau} \left[ \exp\left(2r \|\Gamma_{n,i}(t,y)\|_{T,y^{\dagger}} \right) \right]^{1/2} \cdot \mathbb{E}_{\eta^{*}}^{\pi} \left[ \mathbb{E}_{\xi}^{\tau} \exp\left(2r \|\Gamma_{n,i}^{*}(t,y)\|_{T,y^{\dagger}} \right) \right]^{1/2} \\
= \mathbb{E}_{\eta}^{\tau} \mathbb{E}_{\xi}^{\tau} \left[ \exp\left(2r \|\Gamma_{n,i}(t,y)\|_{T,y^{\dagger}} \right) \right]^{1/2} \left[ \exp\left(2r \|\Gamma_{n,i}(t,y)\|_{T,y^{\dagger}} \right) \right] \\
\leq \mathbb{E}_{\eta}^{\tau} \mathbb{E}_{\xi}^{\tau} \left[ \exp\left(2r \|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^{\dagger}} \right) \right] \\
\leq \mathbb{E}_{\xi}^{\tau} \left[ \exp\left(2r \sum_{(t,y)\in[0,T]\times[0,y^{\dagger}]} \left\{ \max_{1\leq j\leq N_{n}^{*}(y,t)} \left| n^{-1/2} \sum_{i=1}^{j} \xi_{i} \right| \right\} \right) \right] \right]$$
(55)

where the first inequality holds by Lemma 2; the first equality holds because the centering terms cancel out; the second equality holds because  $\xi_i \cdot (\mathbf{1}(\eta_i(\tau_{n,i}) > t) - \mathbf{1}(\eta_i^*(\tau_{n,i}) > t)) \stackrel{d}{=} \mathbf{1}(\eta_i(\tau_{n,i}) > t) - \mathbf{1}(\eta_i^*(\tau_{n,i}) > t))$ ; the second inequality follows by (54) and the triangle inequality; the third equality holds by conditioning on the  $\eta$  and  $\eta^*$ ; the third inequality holds by applying the Cauchy-Schwarz inequality on the expectation  $\mathbb{E}_{\xi}^{\tau}$ ; 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 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 (55), we apply integration by parts as on p. 150 of [8] to write the moment generating function of a non-negative random variable Y as

$$\mathbb{E}\left[e^{\theta Y}\right] = 1 + \int_0^\infty \theta e^{\theta x} \mathbb{P}\left(Y \ge x\right) \, dx.$$
(56)

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 [13] (also Theorem 3 in the appendix); in particular

$$\mathbb{P}^{\tau}\left(\max_{1 \le j \le X_{n}(0)} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{j} \xi_{i} \right| \ge x\right) \le 2\mathbb{P}^{\tau}\left( \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{X_{n}(0)} \xi_{i} \right| \ge x\right) \le 4e^{-\frac{x^{2}}{2X_{n}(0)}},\tag{57}$$

where the *second* inequality follows from Hoeffding's inequality (e.g., Theorem 3.1.3 on p.120 of [13]; also Theorem 4 in the appendix).

Combining (55)–(57) 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

$$\mathbb{E}^{\tau} \left[ \exp\left( r \| \hat{X}_{n,1}^{e,o} \|_{T,y^{\uparrow}} \right) \right] \leq 1 + 4 \int_{0}^{\infty} 2r e^{2rx} e^{-\frac{x^{2}}{2\bar{X}_{n}(0)}} dx$$
  
$$= 1 + 8r \sqrt{2\pi \bar{X}_{n}(0)} e^{2r^{2} \bar{X}_{n}(0)} \int_{0}^{\infty} \left( 2\pi \bar{X}_{n}(0) \right)^{-1/2} e^{-\frac{(x-2r\bar{X}_{n}(0))^{2}}{2\bar{X}_{n}(0)}} dx$$
  
$$= 1 + 8r \sqrt{2\pi \bar{X}_{n}(0)} e^{2r^{2} \bar{X}_{n}(0)} \Phi\left( 2r \sqrt{\bar{X}_{n}(0)} \right) \leq 1 + 8r \sqrt{2\pi \bar{X}_{n}(0)} e^{2r^{2} \bar{X}_{n}(0)},$$
(58)

where  $\Phi$  is the cdf of the standard normal distribution. Here we assume without loss of generality that  $\bar{X}_n(0) > 0$  because (55) becomes 0 if  $\bar{X}_n(0) = 0$ .

Now recall that we are assuming (40), 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

$$\bar{X}_n(0) \le 2X(0) + K \equiv X^{\uparrow} \quad \text{for all} \quad n \ge 1,$$
(59)

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

Letting  $r = 1/\sqrt{\bar{X}_n(0)}$  in (58) and applying (51), we have

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

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

Verifying condition (ii): asymptotically negligible oscillations. We show that the oscillations are asymptotically negligible, again assuming (40). For that purpose, consider an arbitrary sequence of uniformly bounded stopping times  $\{\kappa_n\}$  with respect to the natural filtration  $\mathbf{F}_n \equiv \{\mathcal{F}_n(t), t \in [0, \infty)\} \lor \mathcal{N}$  where

$$\mathcal{F}_{n}(t,y) \equiv \sigma\{1(\eta_{i}(\tau_{n,i}) > x) : 1 \le i \le X_{n}^{e}(0,y), x \ge t, 0 \le s \le y\} \lor \sigma\{X_{n}^{e}(0,(y-x)^{+}), x \ge t, 0 \le s \le y\},$$

$$\mathcal{F}_{n}(t) \equiv \bigvee_{y \ge 0} \mathcal{F}_{n}(t,y),$$

$$(60)$$

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  $\{\kappa_n\}$ ,

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \sup_{\kappa_n} \mathbb{E}^{\tau} \left[ \left( \sup_{y \ge 0} \left| \hat{X}_{n,1}^{e,o}(\kappa_n + \delta, y) - \hat{X}_{n,1}^{e,o}(\kappa_n, y) \right| \right)^2 \right] = 0, \tag{61}$$

which is a sufficient condition for condition (ii) in Theorem 6.2 of [31].

To establish (61), 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  $\{\hat{X}_{n,1}^e(t,y): y \ge 0\}$  with t fixed is an adapted martingale with respect to  $\tilde{\mathbf{F}}_n^t \equiv \bigvee_{y\ge 0} \mathcal{F}_n^t(y) \lor \{\kappa_n\} \lor \{\tau_{n,i}\}$ , where  $\mathcal{F}_n^t(y)$  denotes the  $\sigma$ -algebra  $\mathcal{F}_n(t,y)$  in (60) with t being fixed. Consequently, with the conditioning, the process  $(\hat{X}_{n,1}^{e,o}(\kappa_n + \delta, y) - \hat{X}_{n,1}^{e,o}(\kappa_n, y), y \ge 0)$  is an  $\tilde{\mathbf{F}}_n^{\kappa_n+\delta}$ -adapted martingale. Let  $\mathbb{E}^{\tau,\kappa}$  denote that the expectation is computed by conditioning on  $\{\tau_{n,i}\}$  and  $\{\kappa_n\}$ , let  $\operatorname{Var}^{\tau,\kappa}$  be the conditional variance. Then, by Doob's maximal inequality,

$$\mathbb{E}^{\tau,\kappa} \left( \sup_{y\geq 0} \left| \hat{X}_{n,1}^{e,o}(\kappa_{n}+\delta,y) - \hat{X}_{n,1}^{e,o}(\kappa_{n},y) \right| \right)^{2} \leq 4 \sup_{y\geq 0} \mathbb{E}^{\tau,\kappa} \left( \hat{X}_{n,1}^{e,o}(\kappa_{n}+\delta,y) - \hat{X}_{n,1}^{e,o}(\kappa_{n},y) \right)^{2} \\
= \frac{4}{n} \sup_{y\geq 0} \operatorname{Var}^{\tau,\kappa} \left( \sum_{i=1}^{X_{n}^{e}(0,(y-\kappa_{n})^{+})} \mathbf{1}(\eta_{i}(\tau_{n,i}) > \kappa_{n} + \delta) - \sum_{i=1}^{X_{n}^{e}(0,(y-\kappa_{n})^{+})} \mathbf{1}(\eta_{i}(\tau_{n,i}) > \kappa_{n}) \right) \\
\leq \frac{4}{n} \sup_{y\geq 0} \operatorname{Var}^{\tau,\kappa} \left( \sum_{i=1}^{X_{n}^{e}(0,(y-\kappa_{n})^{+})} \mathbf{1}(\kappa_{n} < \eta_{i}(\tau_{n,i}) \leq \kappa_{n} + \delta) \right) \\
= \frac{4}{n} \sup_{y\geq 0} \sum_{i=1}^{X_{n}^{e}(0,(y-\kappa_{n})^{+})} H^{\delta}_{\tau_{n,i}}(\kappa_{n})(1 - H^{\delta}_{\tau_{n,i}}(\kappa_{n})) \\
= 4 \int_{0}^{\infty} H^{\delta}_{u}(\kappa_{n})(1 - H^{\delta}_{u}(\kappa_{n})) \ d\bar{X}_{n}^{e}(0,u) \\
\leq 4 \int_{0}^{y^{\dagger}} H^{\delta}_{u}(M) \ d\bar{X}_{n}^{e}(0,u) \leq \int_{0}^{y^{\dagger}} \frac{4g_{0}^{\dagger}\delta}{G^{c}(M+y^{\dagger})} \ d\bar{X}_{n}^{e}(0,u) \\
= \frac{4g^{\dagger}\delta}{G^{c}(M+y^{\dagger})} \bar{X}_{n}(0) \leq \frac{4g^{\dagger}\delta}{G^{c}(M+y^{\dagger})} X^{\dagger} \to 0,$$
(62)

as  $\delta \to 0$ . 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 (59). 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. The last inequality uses Assumptions 2 and 3. Therefore, the condition in (61) is satisfied.

4.3. Convergence of the Second Term in (38). In this section, we establish convergence of the second term in (38), i.e.,  $\hat{X}_{n,2}^{e,o} \Rightarrow \hat{X}_2^{e,o}$ , again conditioning on the ages and assuming that (40) holds, so that we can apply Lemma 1. Since  $\hat{X}_n^e(0,\cdot)$  is of bounded variation, the second term in (38) can be expressed as a Stieltjes integral. Therefore, we can use the integration by parts formula given on p.336 of [4] to obtain an equivalent representation

$$\hat{X}_{n,2}^{e,o}(t,y) \equiv -\int_{0}^{(y-t)^{+}} H_{u}^{c}(t)d\hat{X}_{n}^{e}(0,u) = H_{(y-t)^{+}}^{c}(t)\hat{X}_{n}^{e}(0,(y-t)^{+}) - \int_{0}^{(y-t)^{+}} \hat{X}_{n}^{e}(0,u-)dH_{u}^{c}(t).$$
 (63)

Since we are conditioning on the ages, everything in (63) 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  $\sigma$ -field induced by the usual topology on  $\mathbb{D}_{\mathbb{D}}$  coincides with the usual Kolmogorov  $\sigma$ -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. (Finite total variation in x for  $H_x^c(t)$  in bounded intervals) In an interval  $[0, T^*], \int_0^{T^*} |dH_x^c(t)| < \infty$ , for  $t \ge 0$ .

*Proof.* Taking the derivative of  $H_x^c(t)$  with respect to x yields

$$\int_{0}^{T^{*}} |dH_{x}^{c}(t)| = \int_{0}^{T^{*}} \frac{|g(t+x)G^{c}(x) - g(x)G^{c}(t+x)|}{(G^{c}(x))^{2}} dx$$

$$< \left(\frac{g^{\uparrow}}{G^{c}(T^{*})} + \frac{g^{\uparrow}}{(G^{c}(T^{*}))^{2}}\right) T^{*} \equiv K(T^{*}) < \infty. \quad \bullet \quad (64)$$

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

$$\phi(x)(t,y) = H_{(y-t)^+}(t)x((y-t)^+) - \int_0^{(y-t)^+} x(s-)dH_s(t)$$
(65)

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

$$\begin{aligned} \left|\phi(x_n)(t,y) - \phi(x)(t,y)\right| &\leq H_{(y-t)^+}^c(t) \left|x_n((y-t)^+) - x((y-t)^+)\right| + \left|\int_0^{(y-t)^+} (x_n(s-) - x(s-)) dH_s^c(t)\right| \\ &\leq H_{(y-t)^+}^c(t) \|x_n - x\|_{y_0} + \|x_n - x\|_{y_0} \int_0^{(y-t)^+} |dH_s^c(t)| \leq (1 + K(y_0)) \|x_n - x\|_{y_0} \end{aligned}$$

where the finite constant  $K(y_0)$  is defined in (64). Therefore, as  $n \to \infty$ ,

$$d_u(\phi(x_n),\phi(x)) \equiv \sup_{(t,y)\in[0,T]\times[0,\infty)} |\phi(x_n)(t,y) - \phi(x)(t,y)| \to 0.$$
 (66)

Finally, we observe that Lemma 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 (24).

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 [31] 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 (21) and (26). We only prove (21) because (26) is similar. We obtain the right-hand representation in (21) using the fact that  $\{\hat{U}(t,y_0)/\sqrt{y_0(1-y_0)};t\geq 0\}$  is a standard BM motion for a fixed  $0 < y_0 < 1$  (§A of the appendix of [31]). We have

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

which coincides with the right-hand expression in (21). To show that the two expressions in (21) 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 [31], the variance of the first expression in (21) is

$$\begin{split} & \mathbb{E}\left[\left(\int_{0}^{t}\int_{0}^{\infty}\mathbf{1}(x>t-s)\,d\left(\hat{W}(\Lambda(s),G_{\nu}(x))-G_{\nu}(x)\hat{W}(\Lambda(s),1)\right)\right)^{2}\right] \\ &= \mathbb{E}\left[\left(\int_{0}^{t}\int_{0}^{\infty}\mathbf{1}(x>t-s)\,d\hat{W}(\Lambda(s),G_{\nu}(x))-\int_{0}^{t}G_{\nu}^{c}(t-s)\,d\hat{W}(\Lambda(s),1)\right)^{2}\right] \\ &= \int_{0}^{t}\int_{0}^{\infty}\mathbf{1}(x>t-s)\,d\Lambda(s)dG_{\nu}(x)+\int_{0}^{t}G_{\nu}^{c}(t-s)^{2}\,d\Lambda(s)-2\int_{0}^{t}\int_{0}^{\infty}G_{\nu}^{c}(t-s)\mathbf{1}(x>t-s)\,d\Lambda(s)dG_{\nu}(x) \\ &= \int_{0}^{t}G_{\nu}^{c}(t-s)G_{\nu}(t-s)\,d\Lambda(s), \end{split}$$

which simply coincides with the variance of the second expression in (21).

5. The  $G_t/GI^o, GI^\nu/\infty$  Model Starting in the Past We now show that Theorem 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  $G_\nu$  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 6. (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  $G_\nu$  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 2 holds for  $t \ge 0$ , with

$$X^{e}(0,y) = \int_{0}^{y} G^{c}(s)\lambda(-s) \, ds \cdot \mathbf{1}(0 \le y \le t_{0}) + \int_{0}^{y-t_{0}} H^{c}_{x}(t_{0}+y) \, dX^{e}(-t_{0},x) \cdot \mathbf{1}(y > t_{0}), \tag{67}$$

$$\hat{X}^{e}(0,y) = \hat{X}^{e}_{a}(0,y) \cdot \mathbf{1}(0 \le y \le t_{0}) + \left(\hat{X}^{e}_{b,1}(0,y) + \hat{X}^{e}_{b,2}(0,y)\right) \cdot \mathbf{1}(y > t_{0}),$$
(68)

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  $\mathcal{B}_a$ ,

$$\hat{X}_{a}^{e}(0,y) = c_{\lambda} \int_{-y}^{0} G^{c}(-s) d\mathcal{B}_{a}(\Lambda(s)) + \int_{-y}^{0} \int_{0}^{\infty} \mathbf{1}(x > -s) d\hat{K}_{o}(\Lambda(s), x) \\
\stackrel{d}{=}_{y} c_{\lambda} \int_{-y}^{0} G^{c}(-s) d\mathcal{B}_{a}(\Lambda(s)) - \int_{-y}^{0} \sqrt{G(-s)G^{c}(-s)} d\mathcal{B}_{s}(\Lambda(s)) \\
\stackrel{d}{=}_{y} \int_{-y}^{0} \sqrt{(c_{\lambda}^{2} - 1)G^{c}(-s^{2}) + G^{c}(-s)\lambda(s)} d\mathcal{B}(s), \quad for \quad 0 \le y \le t_{0},$$
(69)

 $\hat{X}_{b1}^{e}(0,y)$  is a zero-mean Gaussian process with covariance

$$Cov\left(\hat{X}_{b,1}^{e}(0,y_{1}),\hat{X}_{b,1}^{e}(0,y_{2})\right) = \int_{0}^{y_{1}\wedge y_{2}-t_{0}} H_{u}(t_{0})H_{u}^{c}(t_{0}) dX^{e}(-t_{0},u), \quad for \quad y_{1},y_{2} > t_{0},$$
  
and  $\hat{X}_{b,2}^{e}(0,y) = \int_{0}^{y-t_{0}} H_{x}^{c}(t_{0}) d\hat{X}^{e}(-t_{0},u), \quad for \quad y > t_{0}.$ 

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

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

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

$$\sigma_{\hat{X}}^{2}(t) = \sigma_{\nu}^{2}(-t_{0}, t) = \int_{-t_{0}}^{t} \left[ (c_{\lambda}^{2} - 1)G_{\nu}^{c}(t-s)^{2} + G_{\nu}^{c}(t-s) \right] \lambda(s) \, ds.$$

$$(70)$$

REMARK 9. (Verifying consistency with [31]) Corollary 6 provides an important consistency check by allowing us to compare with the previous results in [31]. In particular, we see that we get strong verification through (70).

We illustrate Corollary 6 with the following example.

EXAMPLE 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

$$\lambda_n(t) = n\left(a + b\sin(ct + \phi)\right), \quad t_0 \le t \le T,\tag{71}$$

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 1 and 2 imply that



FIGURE 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 (71) having parameters a = c = 1, b = 0.6,  $\phi = 0$  and n = 100.

 $X_n(t) \approx nX(t) + \sqrt{n}\hat{X}(t)$  when n is large. Therefore, we expect  $\mathbb{E}[X_n(t)] \approx nX(t)$  and  $\operatorname{Var}(X_n(t)) \approx n\operatorname{Var}(\hat{X}(t))$ . 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 1 and Corollaries 5 and 6. 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 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.

6. 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  $\{\hat{X}_{n}^{e,\nu}(t,\cdot):t\geq 0\}$  in  $\mathbb{D}_{\mathbb{D}}$ , which is itself an element of  $\mathbb{D}$ . As observed by Glynn and Whitt [10], 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 7. (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

$$(\hat{X}_{1}^{e,\nu}(t,y),\hat{X}_{2}^{e,\nu}(t,y)) = \left(\sqrt{\lambda c_{a}^{2}} \int_{(t-y)^{+}}^{t} G^{c}(t-s) \, d\mathcal{B}_{a}(s), \int_{(t-y)^{+}}^{t} \int_{t-s}^{\infty} d\hat{K}(\lambda s,x)\right), \tag{72}$$

$$(\hat{X}_{1}^{r,\nu}(t,y),\hat{X}_{2}^{r,\nu}(t,y)) = \left(\sqrt{\lambda c_{a}^{2}} \int_{-t_{0}}^{t} G^{c}(t+y-s) \, d\mathcal{B}_{a}(s), \int_{-t_{0}}^{t} \int_{t+y-s}^{\infty} d\hat{K}(\lambda s,x)\right). \tag{73}$$

(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  $t + t_0 \ge y^*$  and thus reaches steady state at time  $y^*$  if  $t_0 = 0$  (or is in steady state at time 0 if  $t_0 \ge y^*$ );

(b) As  $-t_0 \downarrow -\infty$ , corresponding to the system starting empty in the distant past, the processes in (72) and (73) converge w.p.1 to the associated stationary processes (as functions of t)

$$(\hat{X}_{1}^{e,\nu}(t,y), \hat{X}_{2}^{e,\nu}(t,y)) = \left(\sqrt{\lambda c_{a}^{2}} \int_{t-y}^{t} G^{c}(t-s) \, d\mathcal{B}_{a}(s), \int_{t-y}^{t} \int_{t-s}^{\infty} d\hat{K}(\lambda s,x)\right), \tag{74}$$

$$(\hat{X}_{1}^{r,\nu}(t,y), \hat{X}_{2}^{r,\nu}(t,y)) = \left(\sqrt{\lambda c_{a}^{2}} \int_{-\infty}^{t} G^{c}(t+y-s) \, d\mathcal{B}_{a}(s), \int_{-\infty}^{t} \int_{t+y-s}^{\infty} d\hat{K}(\lambda s,x)\right), \tag{75}$$

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

$$(\hat{X}_{1}^{e,s}(y), \hat{X}_{2}^{e,s}(y)) \equiv \left(\sqrt{\lambda c_{a}^{2}} \int_{-y}^{0} G^{c}(-s) d\mathcal{B}_{a}(s), \int_{-y}^{0} \int_{-s}^{\infty} d\hat{K}(\lambda s, x)\right), \quad y \ge 0,$$
(76)

$$(\hat{X}_{1}^{r,s}(y), \hat{X}_{2}^{r,s}(y)) \equiv \left(\sqrt{\lambda c_{a}^{2}} \int_{-\infty}^{0} G^{c}(y-s) \, d\mathcal{B}_{a}(s), \int_{-\infty}^{0} \int_{y-s}^{\infty} d\hat{K}(\lambda s, u)\right), \quad y \ge 0, \tag{77}$$

with covariance formulas given by

$$Cov(\hat{X}_{1}^{e,s}(y_{1}), \hat{X}_{1}^{e,s}(y_{2})) = \lambda c_{a}^{2} \int_{0}^{y_{1} \wedge y_{2}} G^{c}(s)^{2} ds, \quad Cov(\hat{X}_{2}^{e,s}(y_{1}), \hat{X}_{2}^{e,s}(y_{2})) = \lambda \int_{0}^{y_{1} \wedge y_{2}} G^{c}(s)G(s) ds,$$

$$Cov(\hat{X}_{1}^{e,s}(y_{1}), \hat{X}_{1}^{e,s}(y_{2})) = \lambda c_{a}^{2} \int_{0}^{\infty} G^{c}(y_{1}+s)G^{c}(y_{2}+s) ds,$$

$$Cov(\hat{X}_{2}^{e,s}(y_{1}), \hat{X}_{2}^{e,s}(y_{2})) = \lambda \int_{0}^{\infty} (G^{c}((y_{1} \vee y_{2})+s) - G^{c}(y_{1}+s)G^{c}(y_{2}+s)) ds.$$
(78)

**Proof** We have already established (72) and (73). 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 (78). We only prove the first two because the proofs of the others are similar. First, by the isometry of Brownian integrals,

$$\operatorname{Cov}(\hat{X}_{1}^{e,s}(y_{1}),\hat{X}_{1}^{e,s*}(y_{2})) = \lambda \int_{-y_{1} \wedge y_{2}}^{0} G^{c}(-s)^{2} \, ds = \lambda \int_{0}^{y_{1} \wedge y_{2}} G^{c}(s)^{2} \, ds.$$
(79)

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 [31]), we have

$$Cov(\hat{X}_{2}^{e,s}(y_{1}), \hat{X}_{2}^{e,s}(y_{2})) = \mathbb{E}\left[\int_{-y_{1}}^{0}\int_{-s}^{\infty}d\left(W(\lambda s, G(x)) - G(x)W(\lambda s, 1)\right) \times \int_{-y_{2}}^{0}\int_{-s}^{\infty}d\left(W(\lambda s, G(x)) - G(x)W(\lambda s, 1)\right)\right] = \lambda \int_{-y_{1}\wedge y_{2}}^{0}G^{c}(-s)ds + \lambda \int_{-y_{1}\wedge y_{2}}^{0}G^{c}(-s)^{2}ds - 2\lambda \int_{-y_{1}\wedge y_{2}}^{0}G^{c}(-s)^{2}ds,$$
(80)

which coincides with the second covariance formula in (78).

Corollary 7 adds to the insight about the stationary model provided by Corollaries 3.1, 4.1 and 4.2 of [31]. As indicated in Corollary 4.1 of [31], we see that the approximating stationary distribution depends on the arrival process beyond its constant arrival rate  $\lambda$  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  $G_e(t) \equiv (1/E[S]) \int_0^t G^c(u) du, t \ge 0$  and ccdf  $G_e^c(t) \equiv 1 - G_e(t)$ ; e.g., see [6, 11]. Thus we have the following corollary.

COROLLARY 8. (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  $t \ge 0$ ) given by

$$X^{e,s}(0,y) = \rho G_e(y) = \int_0^y a(y) dy = \int_0^y \lambda G^c(u) \, du \quad and \quad \hat{X}^{e,s}(0,y) \stackrel{\mathrm{d}}{=} \hat{X}_1^{e,s*}(y) + \hat{X}_2^{e,s*}(y), \quad (81)$$

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

$$\hat{X}_{1}^{e,s*}(y) \equiv \hat{U}^{*}(\rho, G_{e}(y)) \stackrel{d}{=} \sqrt{\rho} \mathcal{B}_{s}^{*}(G_{e}(y)) \quad and \quad \hat{X}_{2}^{e,s*}(y) \equiv G_{e}(y) \hat{X}(0) \stackrel{d}{=} \sqrt{\rho} G_{e}(y) \mathcal{Z}_{0}, \tag{82}$$

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

$$X^{r,s}(0,y) = \rho G_e^c(y) = \int_y^\infty \lambda G^c(u) \, du \quad and \quad \hat{X}^{r,s}(0,y) \stackrel{\mathrm{d}}{=} \hat{X}_1^{r,s*}(y) + \hat{X}_2^{r,s*}(y), \quad t \ge 0, \tag{83}$$

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

$$\operatorname{Cov}\left(\hat{X}_{1}^{r,s*}(x_{1}), \hat{X}_{1}^{r,s*}(x_{2})\right) = \int_{0}^{\infty} H_{u}(x_{1} \wedge x_{2}) H_{u}^{c}(x_{1} \vee x_{2}) \, dX^{e,s}(0,u),$$
  
and  $\hat{X}_{2}^{r,s*}(x) \stackrel{\mathrm{d}}{=} \sqrt{\rho} \int_{0}^{\infty} H_{u}^{c}(x) \, d\mathcal{B}_{o}^{*}(G_{e}(u)) + \sqrt{\rho} G_{e}^{c}(y) \, \mathcal{Z}_{0}.$  (84)

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

$$\operatorname{Var}(\hat{X}^{e,s*}(y)) = \int_0^y \lambda G^c(u) \, du, \quad y \ge 0.$$
(85)

As a consequence, the variance of the total content, as the sum of variances of the new content (that is (33)) and old content (that is (34)), is

$$\sigma_{\hat{X}}^2 = \sigma_{\hat{X},\nu}^2 + \sigma_{\hat{X},o}^2 = \lambda \int_0^t G^c(u) du + \lambda \int_0^\infty H_u^c(t) G^c(u) du = \lambda \int_0^\infty G^c(u) du = \rho.$$
(86)

**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 (81), we have

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

where the convergence holds because the age  $A_i$  follows the equilibrium distribution  $G_e$  (see Theorem 1 of [6]). To prove the FCLT in (81), we have

$$\hat{X}^{e}(0,y) = \frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n\bar{X}_{n}(0)} \mathbf{1}(A_{i} \leq y) - nX(0) G_{e}(y) \right) \\
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n\bar{X}_{n}(0)} \left( \mathbf{1}(A_{i} \leq y) - G_{e}(y) \right) + \hat{X}_{n}(0) G_{e}(y) = \hat{K}_{n} \left( \bar{X}_{n}(0), G_{e}(y) \right) + \hat{X}_{n}(0) G_{e}(y) \quad (87) \\
\Rightarrow \hat{U}^{*} \left( X(0), G_{e}(y) \right) + \hat{X}(0) G_{e}(y) \stackrel{d}{=} \sqrt{X(0)} \mathcal{B}^{*}_{s} \left( G_{e}(y) \right) + \sqrt{X(0)} G_{e}(y) \mathcal{Z}_{0} \quad \text{in } \mathbb{D},$$

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 (87) (due to the Gaussian CLT limit for a Poisson random variable) and by (ii) Lemma 1 and the conditional independence of these two terms conditioning on  $X_n(0)$  (with arguments similar to §4.1). The variance formulas (85) and (86) immediately follow from (83) and (34) with  $G_{\nu} = G$ . The proof of the FWLLN limit in (83) follows from (81) and Theorem 1, and the proof of (84) follows from (83), Theorem 2 and Corollary 3.

COROLLARY 9. (Two equivalent decompositions) For the  $M/GI/\infty$  model (with  $c_a^2 = 1$ ), the two representations (76) and (81) ((77) and (83)) are equivalent independent decompositions, i.e.,

$$\hat{X}_{1}^{e,s*} + \hat{X}_{2}^{e,s*} \stackrel{\mathrm{d}}{=} \hat{X}_{1}^{e,s} + \hat{X}_{2}^{e,s} \quad and \quad \hat{X}_{1}^{r,s*} + \hat{X}_{2}^{r,s*} \stackrel{\mathrm{d}}{=} \hat{X}_{1}^{r,s} + \hat{X}_{2}^{r,s}.$$
(88)

**Proof** We only prove the first equality in (88) since the second equality follows similarly. Because all four terms in the first equality of (88) 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

$$\sum_{k=1}^{2} \operatorname{Cov}(\hat{X}_{k}^{e,s*}(y_{1}), \hat{X}_{k}^{e,s*}(y_{2})) = \sum_{k=1}^{2} \operatorname{Cov}(\hat{X}_{k}^{e,s}(y_{1}), \hat{X}_{k}^{e,s}(y_{2})),$$
(89)

By (82), we have

$$\begin{aligned} &\operatorname{Cov}(X_1^{e,s*}(y_1), X_1^{e,s*}(y_2)) = \rho \left[ G_e(y_1) \wedge G_e(y_2) - G_e(y_1) G_e(y_2) \right], \\ &\operatorname{Cov}(\hat{X}_2^{e,s*}(y_1), \hat{X}_2^{e,s*}(y_2)) = \rho G_e(y_1) G_e(y_2). \end{aligned}$$

so that the left-hand side of (89) 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 (89), according to (78) 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, even though the steady-state distribution at one time is remarkably simple. We now illustrate with an example.

EXAMPLE 2. (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  $\mu_1$  and  $\mu_2$  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 [6]. See Figure 2 for the simulation comparison, which is an analog of Figure 1.



FIGURE 2. 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  $n\lambda = 100$ .

7. 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, and exploit

COROLLARY 10. (Initial customers generated from a renewal process with a time change) In the n<sup>th</sup>  $G_t/GI/\infty$  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 interrenewal SCV  $\bar{c}_0^2$  and the deterministic function

$$X^*(y) \equiv \int_0^y b^*(y) dy < \infty, \tag{90}$$

then Assumption 2 is satisfied with  $X(0,y) = X^*(y)$  and  $\hat{X}(0,y) = \bar{c}_0 \mathcal{B}_0(X^*(y))$ , for  $y \ge 0$ , where  $\mathcal{B}_0$  is a standard BM. In addition, the variance of the ICP (that is (34)), is

$$\sigma_{\hat{X},o}^{2}(t) = \int_{0}^{\infty} H_{u}(t)H_{u}^{c}(t) \, dX^{*}(u) + \bar{c}_{0}^{2} \int_{0}^{\infty} H_{u}^{c}(t)^{2} \, dX^{*}(u) = \int_{0}^{\infty} \left[ \left( \bar{c}_{0}^{2} - 1 \right) H_{u}^{c}(t) + 1 \right] H_{u}^{c}(t) b^{*}(u) \, du.$$

EXAMPLE 3. (Simulation comparison for an example of Corollary 10) 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 (71) 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 10 and consider two density functions in (90)

$$b_1^*(u) = u \mathbf{1}_{(0 \le u \le d)} + (2d - u) \mathbf{1}_{(d \le u \le 2d)} \quad \text{and} \quad b_2^*(u) = \frac{1}{3} u^2 \mathbf{1}_{(0 \le u \le 2d)}, \tag{91}$$

with d = 1.5, as shown in Figure 3.



FIGURE 3. Two choices of the initial-condition density  $b^*(u)$  in (90).

Comparisons with simulations are shown in Figure 4. The first three plots in Figure 4 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.

8. Conclusion Our main result is an FCLT for the initial content process (ICP), which specifies the elapsed service times of all customers in the system at time t who were also initially in the system at time 0, having arrived sometime earlier. The ICP in system n is the process  $X_n^{e,o}(t, y)$  in  $\mathbb{D}_{\mathbb{D}}$ , defined in (3). The key assumptions are (i) that the service times come from a sequence of i.i.d. random variables independent of the arrival process and (ii) an FCLT holds for the initial ICP at time 0 in  $\mathbb{D}$ , which includes having the total number at time 0 grow proportional to n. The limit process for the scaled ICP is  $\hat{X}^{e,o}(t,y) \equiv \hat{X}_1^{e,o}(t,y) + \hat{X}_2^{e,o}(t,y)$ , where  $\hat{X}_1^{e,o}(t,y)$  and  $\hat{X}_2^{e,o}(t,y)$  are independent zero-mean Gaussian processes specified in (17) and (18). The proof of the FCLT in §4 exploits results in [37] for empirical processes of independent non-identically-distributed random variables.

In Theorem 2, our FCLT for the ICP is expressed in the more general context of an  $G_t/GI/\infty$  infinite-server queueing model, which includes new input as well as the ICP. In this context, the



FIGURE 4. Example 3 with  $b_1^*$  in (91): 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 (71) having parameters a = c = 1, b = 0.6,  $\phi = 0$  and n = 100 and general initial conditions.

key assumption is Assumption 1, which requies a joint FCLT for the arrival process and the ICP at time 0, where the limit processes are independent zero-mean Gaussian processes. The limit for the ICP alone is obtained by having a null arrival process. The additional results for the new input are obtained from [31]. Corollary 6 shows that Theorem 2 implies a corresponding FCLT for the key processes in the more general  $G_t/GI^o, GI^\nu/\infty$  model starting at time  $-t_0 < 0$ , which has one service-time cdf G before time 0 and then another service-time cdf  $G_\nu$  after time 0. The limit holds if Assumption 1 holds at time  $-t_0$ , provided that the limit process is a time-transformed Brownian motion, which is the common case for applications.

As a further consequence, in Corollary 7 we characterize the steady-state new and old content in the stationary  $G/GI/\infty$  model. Using the explicit structure of the  $M/GI/\infty$  model, we obtain an alternative characterization of the steady-state new and old content in Corollary 8. We then show that the two different representations are equivalent in Corollary 9. In §5-§7 we also report results of simulation experiments that verify the accuracy of the formulas and show that they can be useful engineering approximations.

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

References

- [1] A. K. Aras and Y. Liu, Functional weak law of large numbers for the time-varying  $(G_t/GI/s_t+GI)^m/M_t$  queueing network. Working paper, 2014.
- [2] A. K. Aras, Y. Liu and W. Whitt, Many-server heavy-traffic limits for the time-varying  $G_t/GI/s_t + GI$  queue. In preparation, 2014.
- [3] P. Billingsley, Convergence of Probability Measures, second ed., Wiley, New York, 1999.
- [4] P. Bremaud, Point Processes and Queues: Martingale Dynamics, Springer-Verlag, New York, 1981.
- [5] L. Decreusefond and P. Moyal, A functional central limit theorem for the M/GI/∞ queue. Ann. Appl. Probab. 18(6) (2008) 2156-2178.
- [6] N. Duffield and W. Whitt, Control and recovery from rare congestion events in a large multi-server system. Queueing Systems 26 (1997) 69–104.
- [7] S. N. Ethier and T. G. Kurtz, Markov Processes: Characterization and Convergence, Wiley, New York, 1986.
- [8] W. Feller, An Introduction to Probability Theory and its Applications, John Wiley and Sons, New York, 1971.
- [9] I. Gerhardt and B. Nelson, Transforming renewal processes for simulation of nonstationary arrival processes. *INFORMS J. on Computing* 21 (2009) 630–640.
- [10] P. W. Glynn and W. Whitt, A new view of the heavy-traffic limit theorem for the infinite-server queue. Adv. Appl. Probab. 23 (1991) 188–209.
- [11] D. Goldberg and W. Whitt, The last departure time from an  $M_t/G/\infty$  queue with a terminating arrival process. Queueing Systems 58 (2008) 77–104.
- [12] L. V. Green, P. J. Kolesar and W. Whitt, Coping with time-varying demand when setting staffing requirements for a service system. *Production and Operations Management* 16 (2007) 13–39.
- [13] A. Gut, Probability: A Graduate Course. Springer, 2005.
- [14] J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*. Springer (1987)
- [15] W. Kang and G. Pang, Equivalence of Fluid Models for  $G_t/GI/N + GI$  Queues. Working paper, 2014.
- [16] W. Kang and K. Ramanan, Fluid limits of many-server queues with reneging. Ann. Appl. Prob. 20 (2010) 2204–2260.
- [17] H. Kaspi and K. Ramanan, Law of large numbers limits for many-server queues. Ann. Appl. Prob. 21 (2011) 33–114.
- [18] H. Kaspi and H., K. Ramanan, SPDE limits of many-server queues. Ann. Appl. Prob. 23 (2013) 149-229.
- [19] E. V. Krichagina and A. A. Puhalskii, A heavy-traffic analysis of a closed queueing system with a GI/∞ service center. Queueing Systems 25 (1997) 235–280.
- [20] A. Li and W. Whitt, Approximate blocking probabilities for loss models with independence and distribution assumptions relaxed. *Performance Evaluation*, published online, August 13, 2013, (http://www.sciencedirect.com/science/article/pii/S0166531613000825)
- [21] R. Liu, Y. Liu and J. R. Wilson, Modeling and simulation of nonstationary and non-Poisson processes. Working paper, North Carolina State University, 2014.
- [22] Y. Liu and W. Whitt, Stabilizing customer abandonment in many-server queues with time-varying arrivals. Oper. Res. 60(6) (2012) 1551–1564.
- [23] Y. Liu and W. Whitt, A many-server fluid limit for the  $G_t/GI/s_t + GI$  queueing model experiencing periods of overloading. Oper. Res. Letters 40 (2012) 307–312
- [24] Y. Liu and W. Whitt, Many-server heavy-traffic limit for queues with time-varying parameters. Ann. Appl. Probab. 24 (2014) 378–421
- [25] Y. Liu and W. Whitt, Stabilizing performance in networks of queues with time-varying arrival rates. Prob. in the Engr. and Inf. Sciences. Forthcoming (2014).
- [26] G. Louchard, Large finite population queueing systems. Part 1: the infinite server model Comm. Statist. Stochastic Models 4 (1988) 473–505.

- [27] A. Mandelbaum and P. Momcilovic, Queues with many servers and impatient customers. Math. Oper. Res. 37(1) (2012) 41–65.
- [28] M. B. Marcus and J. Zinn, The bounded law of iterated logarithm for the weighted empirical distribution process in the non-i.i.d case. Ann. Prob. 12(2) (1984) 335–360.
- [29] W. A. Massey and W. Whitt, Unstable asymptotics for nonstationary queues. Math. Oper. Res. 19 (1994) 267–291.
- [30] G. Pang, R. Talreja and W. Whitt, Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probability Surveys* 4 (2007) 193–267.
- [31] G. Pang and W. Whitt, Two-parameter heavy-traffic limits for infinite-server queues. Queueing Systems 65 (2010) 325–364.
- [32] G. Pang and W. Whitt, The impact of dependent service times on large-scale service systems. Manufacturing and Service Operations Management 14(2) (2012) 262–278.
- [33] G. Pang and W. Whitt, Two-parameter heavy-traffic limits for infinite-server queues with dependent service times. *Queueing Systems* 73 (2013) 119–146.
- [34] A. A. Puhalskii and J. Reed, On many-server queues in heavy traffic. Ann. Appl. Prob. 20 (2010) 129–195.
- [35] J. E. Reed, The G/GI/N queue in the Halfin-Whitt regime. Ann. Appl. Prob. 19 (2009) 2211–2269.
- [36] J. E. Reed and R. Talreja, Distribution-valued heavy-traffic limits for  $G/GI/\infty$  queues. Ann. Appl. Prob. Forthcoming. (2014)
- [37] G. R. Shorack and J. A. Wellner, Empirical Processes with Applications to Statistics. SIAM, 2009.
- [38] A. V. Skorohod, Limit theorems for stochastic processes. Theor. Probability Appl. 1 (1956) 261–290.
- [39] R. Talreja and W. Whitt, Heavy-traffic limits for waiting times in many-server queues with abandonment. Ann. Appl. Prob. 19 (2009) 2137–2175.
- [40] W. Whitt, On the heavy-traffic limit theorem for  $GI/G/\infty$  queues. Adv. Appl. Probab. 14 (1982) 171–190.
- [41] W. Whitt, *Stochastic-Process Limits*, Springer, New York, 2002.
- [42] W. Whitt, Offered load analysis for staffing. Manufacturing and Service Operations Management 15(2) 166–169.
- [43] J. Zhang, Fluid models of many-server queues with abandonment. Queueing Systems. 73 (2013) 147–193.

## Appendix

A. Overview This appendix has extra materials to supplement the main paper. In §B we review useful results used in the proofs. In §C we give proofs omitted in the main paper. In §D we provide additional simulation results.

**B. Useful Inequalities** In this section we review two useful inequalities. Both are used to prove (57) in §4.

THEOREM 3 (Lévy's inequality (symmetric case), Theorem 3.7.1 of Gut[13]). Let  $X_1$ ,  $X_2, \ldots, X_n$  be independent real-valued symmetric random variables (satisfying  $-X_i \stackrel{d}{=} X_i$  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,

$$\mathbb{P}\left(\max_{1\le k\le n} |S_k| > x\right) \le 2\mathbb{P}(|S_n| > x).$$
(92)

THEOREM 4 (Hoeffding's inequality, Theorem 3.1.3 of Gut[13]). Let  $X_1, X_2, ..., X_n$  be independent real-valued random variables such that  $\mathbb{P}(a_k \leq X_k \leq b_k) = 1$  for  $a_k, b_k \in \mathbb{R}$ , k = 1, ..., n, and let  $S_n = \sum_{k=1}^n X_k$ ,  $n \geq 1$ , denote the partial sums. Then

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| > x) \le 2 \exp\left(-\frac{2x^2}{\sum_{k=1}^n (b_k - a_k)^2}\right).$$
(93)

**C. Additional Proofs** We now provide proofs of Corollaries 4 and 6, and Lemma 2, which were omitted in the main paper.

**Proof of Corollary 4.** This follows from parts (ii) and (iii) of Theorem 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

$$\begin{aligned} \operatorname{Cov}(\hat{X}_{1}^{e,\nu}(t_{1},y_{1}),\hat{X}_{1}^{e,\nu}(t_{2},y_{2})) &= \mathbb{E}\left[\int_{(t_{1}-y_{1})^{+}}^{t_{1}} G_{\nu}^{c}(t_{1}-s) \, d\hat{N}(s) \times \int_{(t_{2}-y_{2})^{+}}^{t_{2}} G_{\nu}^{c}(t_{2}-s) \, d\hat{N}(s)\right] \\ &= c_{\lambda}^{2} \int_{(t_{1}-y_{1})^{+} \vee (t_{2}-y_{2})^{+}}^{t_{1} \wedge t_{2}} G_{\nu}^{c}(t_{1}-s) G_{\nu}^{c}(t_{1}-s) \, d\Lambda(s),\end{aligned}$$

and

$$\begin{aligned} \operatorname{Cov}(\hat{X}_{2}^{e,o}(t_{1},y_{1}),\hat{X}_{2}^{e,o}(t_{2},y_{2})) &= \mathbb{E}\left[\int_{0}^{(y_{1}-t_{1})^{+}}H_{x}^{c}(t_{1})\,d\hat{X}_{0}(x) \times \int_{0}^{(y_{2}-t_{2})^{+}}H_{x}^{c}(t_{2})\,d\hat{X}_{0}(x)\right] \\ &= \int_{0}^{(y_{1}-t_{1})^{+}\wedge(y_{2}-t_{2})^{+}}H_{x}^{c}(t_{1})H_{x}^{c}(t_{2})\,d\Sigma_{2}^{e,o}(x).\end{aligned}$$

Second, the covariance of  $X_2^{\nu}$ 

$$\begin{aligned} &\operatorname{Cov}(\hat{X}_{2}^{e,\nu}(t_{1},y_{1}),\hat{X}_{2}^{e,\nu}(t_{2},y_{2})) \\ &= \mathbb{E}\left[\left(\int_{(t_{1}-y_{1})^{+}}^{t_{1}}\int_{0}^{\infty}\mathbf{1}_{(x+s>t_{1})}\,d\left(\hat{W}(\Lambda(s),G_{\nu}(x))-G_{\nu}(x)\hat{W}(\Lambda(s),1)\right)\right)\right) \\ &\times\left(\int_{(t_{2}-y_{2})^{+}}^{t_{2}}\int_{0}^{\infty}\mathbf{1}_{(x+s>t_{2})}\,d\left(\hat{W}(\Lambda(s),G_{\nu}(x))-G_{\nu}(x)\hat{W}(\Lambda(s),1)\right)\right)\right] \\ &= \mathbb{E}\left[\int_{(t_{1}-y_{1})^{+}}^{t_{1}}\int_{0}^{\infty}\mathbf{1}_{(x+s>t_{1})}\,d\hat{W}(\Lambda(s),G_{\nu}(x))\times\int_{(t_{2}-y_{2})^{+}}^{t_{2}}\int_{0}^{\infty}\mathbf{1}_{(x+s>t_{2})}\,d\hat{W}(\Lambda(s),G_{\nu}(x))\right] \end{aligned}$$

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

**Proof of Corollary 6.** First, (67) and (68) easily follow Theorems 3.1 and 3.2 of [31] 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 (68) and Theorem 4.2 of [31], the variance function of  $\hat{X}^e(0, y)$  is

$$\operatorname{Var}(\hat{X}^{e}(0,y)) = \int_{0}^{y} \left[ (c_{\lambda}^{2} - 1)G^{c}(s)^{2} + G^{c}(s) \right] \lambda(-s) \, ds, \quad \text{for} \quad 0 \le y \le t_{0}.$$
(94)

Now plugging (67) and (94) into (34) yields that

$$\begin{split} \sigma_{\hat{X},o}^{2}(t) &= \int_{0}^{t_{0}} H_{u}(t) H_{u}^{c}(t) G^{c}(u) \lambda(-u) \, du + \int_{0}^{t_{0}} \left( H_{u}^{c}(t) \right)^{2} \left[ (c_{\lambda}^{2} - 1) G^{c}(u)^{2} + G^{c}(u) \right] \lambda(-u) \, du \\ &= \int_{0}^{t_{0}} H_{u}(t) G^{c}(t+u) \lambda(-u) \, du + \int_{0}^{t_{0}} H_{u}^{c}(t) G^{c}(t+u) \left[ (c_{\lambda}^{2} - 1) G^{c}(u) + 1 \right] \lambda(-u) \, du \\ &= \int_{0}^{t_{0}} G^{c}(t+u) \left[ (c_{\lambda}^{2} - 1) G^{c}(t+u) + 1 \right] \lambda(-u) \, du \\ &= \int_{-t_{0}}^{0} G^{c}(t-s) \left[ (c_{\lambda}^{2} - 1) G^{c}(t-s) + 1 \right] \lambda(s) \, ds, \end{split}$$

where the last equality holds by a change of variable. Summing the above equation with  $\sigma_{\hat{X},\nu}^2(t)$  in (33) yields (70).

**Proof of Lemma 2.** We mimic the proof of (A.14.15) in [37]. First for a deterministic  $g \in \mathbb{D}_{\mathbb{D}}$ , we have

$$\mathbb{E}||g - X||_{T,y^{\uparrow}} \ge \mathbb{E}|g(t,y) - X(t,y)| \ge |g(t,y) - \mathbb{E}[X(t,y)]|, \quad \text{for} \quad t \in [0,T], \ y \in [0,y^{\uparrow}],$$

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

$$\mathbb{E}\|g - X\|_{T,y^{\uparrow}} \ge \sup_{(t,y) \in [0,T] \times [0,y^{\uparrow}]} |g(t,y) - \mathbb{E}[X(t,y)]| = \|g - \mathbb{E}[X]\|_{T,y^{\uparrow}}.$$
(95)

By conditioning on X, we have

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

where the first inequality holds by Jensen's inequality and the second inequality holds by (95).

## D. An Additional Example

EXAMPLE 4. (Simulation comparison for another example of Corollary 10) 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 (91). Comparisons with simulations are shown in Figure 5.



FIGURE 5. Example 4 with  $b_2^*$  in (91): 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 (71) having parameters a = c = 1, b = 0.6,  $\phi = 0$  and n = 100 and general initial conditions.