The Fluid Limit of an Overloaded Processor Sharing Queue

Amber L. Puhahttp://www.csusm.edu/puha/, Alexander L. Stolyarhttp://cm.bell-labs.com/cm/ms/who/stolyar/, Ruth J. Williamshttp://www.math.ucsd.edu/~williams/,

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 31, No. 2, May 2006, pp. 316–350
ISSN 0364-765X | EISSN 1526-5471 | 06 | 3102 | 0316

DOI 10.1287/moor.1050.0181

The Fluid Limit of an Overloaded Processor Sharing Queue

Amber L. Puha
Department of Mathematics, California State University, San Marcos, 333 Twin Oaks Valley Road,
San Marcos, California 92096–0001, USA, apuha@csusm.edu, http://www.csusm.edu/puha/

Alexander L. Stolyar
Bell Laboratories Lucent Technologies, 600 Mountain Avenue, Room 2C-322, Murray Hill, New Jersey 07974–0636, USA,
stolyar@research.bell-labs.com, http://cm.bell-labs.com/cm/ms/who/stolyar/

Ruth J. Williams
Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0112, USA,
williams@math.ucsd.edu, http://www.math.ucsd.edu/˜williams/

This paper primarily concerns *strictly supercritical fluid models*, which arise as functional law of large numbers approximations for overloaded processor sharing queues. Analogous results for *critical fluid models* associated with heavily loaded processor sharing queues are contained in Gromoll et al. [9] and Puha and Williams [15]. An important distinction between critical and strictly supercritical fluid models is that the total mass for a solution that starts from zero grows with time for the latter, but it is identically equal to zero for the former. For strictly supercritical fluid models, this paper contains descriptions of each of the following: the distribution of the mass as it builds up from zero, the set of stationary solutions, and the limiting behavior of an arbitrary solution as time tends to infinity. In addition, a *fluid limit result* is proved that justifies strictly supercritical fluid models as first order approximations to overloaded processor sharing queues.

Key words: overloaded processor sharing queue; supercritical fluid models; measure-valued process; invariant shape; order preservation; continuity in initial conditions; renewal equations
MSC2000 subject classification: Primary: 60K25; secondary: 68M20, 90B22
OR/MS subject classification: Primary: queues-approximations, queues-limit theorems, queues-transient results; secondary: probability-stochastic model applications
History: Received August 13, 2004; revised April 4, 2005.

1. Introduction. The main aim of this paper is to study strictly supercritical fluid models which arise as functional law of large numbers approximations for overloaded GI/GI/1 processor sharing queues. An analogous study of critical fluid models associated with heavily loaded processor sharing queues is contained in Gromoll et al. [9] and Puha and Williams [15], and the investigation undertaken here is a natural progression from that work. The introduction of Gromoll et al. [9] contains a brief summary of the processor sharing literature, and we refer the reader to that paper for a list of references. The queueing model setup used here is similar to that in Gromoll et al. [9]. In particular, we use a *measure-valued state descriptor* in order to capture information about residual service times. Measure-valued state descriptors have been successfully used for other stochastic networks models; see, e.g., Doytchinov et al. [3], Grishechkin [7], and Gromoll [8]. In formulating the *strictly supercritical fluid models*, we take formal limits of sequences of *measure-valued stochastic processes*, corresponding to sequences of overloaded processor sharing queues, under law of large numbers scaling. Solutions of the strictly supercritical fluid models are functions of time that take values in the set of finite, nonnegative Borel measures on the nonnegative real numbers and that satisfy certain properties. There is an important distinction between the *critical fluid models* (associated with sequences of heavily loaded processor sharing queues) and the strictly supercritical fluid models. Namely, the total mass for a fluid model solution with zero initial measure grows with time for the latter, but it is identically equal to zero for the former. In analyzing the strictly supercritical fluid models, we describe the distribution of mass as it builds up from zero. The same analysis produces some new results for critical fluid models, which we include here as well. Thus, we consider *supercritical fluid models* in this paper. These include both the critical and the strictly supercritical cases. However, the main emphasis is on the latter.

For the strictly supercritical fluid models, we prove four results. We give conditions for existence and uniqueness of fluid model solutions (cf. Theorem 3.1). We explicitly identify the fluid model solutions having a zero initial condition (cf. Theorem 3.2). We identify the fluid model solutions for which the shape does not change with time, the so-called stationary fluid model solutions (cf. Theorem 3.6). Finally, we investigate the asymptotic behavior of fluid model solutions as time tends to infinity. Indeed, under mild conditions, we prove a limit theorem that identifies the asymptotic shape (cf. Theorem 3.5). There are two key elements in the proofs of these theorems. The first is an order preservation property for fluid model solutions (cf. Theorem 4.2), and the second is a continuity property (cf. Theorem 4.3). Both properties are valid not only for strictly supercritical fluid models, but also for critical fluid models. As a consequence, we are able to extend the work in Gromoll et al. [9] to prove a result that holds for both critical and strictly supercritical fluid models (cf. Theorem 3.3). This result states that the mapping from initial conditions in a certain domain to the associated fluid model solutions is continuous.

In order to justify the strictly supercritical fluid models as first-order approximations to overloaded processor sharing queues, we prove a fluid limit result. For this, we consider a sequence of measure-valued stochastic processes corresponding to a sequence of overloaded processor sharing queues. The fluid limit result states that, under mild conditions, the fluid scaled measure-valued stochastic processes converge in distribution to a limit which is almost surely a fluid model solution (cf. Theorem 3.7). This result concerns a sequence of stochastic processes that converges in distribution to a limiting stochastic process. However, since the limit process is almost surely a fluid model solution, it is only stochastic through the initial condition.

In [10], Jean-Marie and Robert prove a law of large numbers type of result that is consistent with our fluid limit result. They study the asymptotic behavior as time tends to infinity of a single overloaded processor sharing queue that has exactly one job in the system at time zero. They prove that, under mild conditions, the state at time t divided by t converges almost surely as t tends to infinity to what we would call the state of the fluid model solution at time one when the initial measure is the zero measure (cf. Jean-Marie and Robert [10, Proposition 4] and the last paragraph of §3.3.2 here).

Fluid approximations for the queue length processes of processor sharing queues are proposed in Chen et al. [2]. As was noted in the introduction of Gromoll et al. [9], the justification given in [2] for those fluid approximations is valid, provided that all limit points are deterministic. Our approach provides a proof that this assumption is valid for the overloaded processor sharing queue under the mild conditions identified here (cf. Theorem 3.7). In the strictly supercritical setting, the authors of Chen et al. [2] show that their fluid approximations for the queue length processes are asymptotically linear as time tends to infinity, and they determine the limiting linear rates of increase (cf. [2, Proposition 6] and the last paragraph in §3.2.7 here). Our Theorem 3.5 is a generalized version of [2, Proposition 6] that also specifies the asymptotic shape (as time goes to infinity) of the fluid approximation to the measure-valued process that captures the behavior of the residual service times.

In [13], our fluid limit result is used by Mandjes and Zwart who analyze the asymptotic behavior of the tail probability for the sojourn time of a tagged job in a strictly subcritical GI/GI/1/PS queue in steady state. The main Mandjes and Zwart result [13, Theorem 3.1] concerns the sojourn time under a light-tailed (but not too light) service time distribution (cf. [13, Assumptions 3.1 and 3.2]). In the proof of Mandjes and Zwart [13, Theorem 3.1], the tail probability for the sojourn time in a strictly subcritical queue is related to the tail probability for the sojourn time in a strictly supercritical queue (cf. [13, Equation (3.7)]), and then the strictly supercritical tail probability in [13, Equation (3.7)] is analyzed to obtain a desired lower bound. The analysis of the strictly supercritical tail probability in [13, Equation (3.7)] in part relies on [13, Lemma 3.1], the proof of which uses our fluid limit result.

The paper is organized in the following manner. Section 2 describes the basic setup. Specifically, §2.1 establishes the notation that is used throughout the paper, and §2.2 describes the dynamics for a processor sharing queue and the associated measure-valued state descriptor. Section 3 contains a discussion of the fluid models. The definitions of fluid model solutions for both the critical and strictly supercritical models are contained in §3.1. Section 3.2 contains the statements of the main results proved in the paper concerning solutions of the supercritical fluid models (Theorems 3.1, 3.2, 3.5, and 3.6 for strictly supercritical data, and Theorem 3.3 for supercritical data). Section 3.3 contains the statement of the fluid limit result (Theorem 3.7). Sections 4 and 5 are devoted to the proofs of the theorems stated in §3. Some of the theorems can be proved in a very similar manner to those in Gromoll et al. [9], particularly for the case of nonzero initial measures. For these instances, we have chosen to omit the complete proof in favor of referencing the similar proof in Gromoll et al. [9] and explaining how to adapt it. This allows us to focus on the new tools and results such as the order preservation property (Lemma 4.2 in §4.3), the proof of uniqueness for Theorem 3.2, the proof of continuity at the zero initial measure for Theorem 3.3, and establishing tightness for the fluid limit result.

2. Preliminaries and queueing model.

In this section, we outline some necessary background. In §2.1, we introduce the basic notation that is used throughout the paper. In §2.2, we describe the processor sharing queueing model. Readers who are familiar with the setup in Gromoll et al. [9] will find this to be rather redundant, since it is largely taken verbatim from that paper. We repeat it here for completeness because the processor sharing queueing model needs to be precisely defined in order to state the fluid limit result.

2.1. Notation.

Let $\mathbb{N}$ denote the set of natural numbers, let $\mathbb{R}$ denote the set of real numbers, and let $\mathbb{R}_+$ denote the set of nonnegative real numbers. For $a, b \in \mathbb{R}$, we write $a \vee b$ for the maximum of a and b, $a \wedge b$

for the minimum of a and b, $a^+ = a \wedge 0$ and $a^- = a \vee 0$ for the positive and negative parts of a respectively, $\lfloor a \rfloor$ for the largest integer less than or equal to a, and $\lceil a \rceil$ for the smallest integer greater than or equal to a. For a Borel-measurable function g: $\mathbb{R}_+ \to \mathbb{R}$, define the positive and negative parts of g by $g^+(x) = g(x) \vee 0$ and $g^-(x) = (-g(x)) \vee 0$ for all $x \in \mathbb{R}_+$. For such a function g, let $\|g\|_\infty = \sup_{x\in\mathbb{R}_+} |g(x)|$, $\|g\|_K = \sup_{x\in[0,K]} |g(x)|$ for each $K \geq 0$, and $\|g\|_{L^1} = \int_0^\infty |g(x)|\,dx$, where we allow these quantities to take the value $+\infty$.

For a set $B \subset \mathbb{R}_+$, we denote the indicator function for the set B by 1_B. We also define the real-valued functions χ and φ on $\mathbb{R}_+$ as follows: $\chi(x) = x$ for $x \in \mathbb{R}_+$, and $\varphi(x) = 1/x$ for $x \in (0,\infty)$ and $\varphi(0) = 0$. For a topological space A, denote by $\mathbf{C}(A)$ the set of continuous, real-valued functions defined on A. Similarly, $\mathbf{C}_b(A)$ denotes the set of continuous, bounded, real-valued functions defined on A. In addition, for an interval $I \subset \mathbb{R}$, $\mathbf{C}^1(I)$ is the set of once continuously differentiable, real-valued functions defined on I, and $\mathbf{C}^1_b(I)$ is the set of functions in $\mathbf{C}^1(I)$ that together with their first derivatives are bounded on I. For $g \in \mathbf{C}^1(I)$, we write $g'(x) = (\frac{d}{dx})g(x)$, for all $x \in I$.

Let $\mathcal{M}_F$ be the set of finite, nonnegative Borel measures on $\mathbb{R}_+$. Consider $\zeta \in \mathcal{M}_F$ and a Borel-measurable function g: $\mathbb{R}_+ \to \mathbb{R}$ which is integrable with respect to ζ. We define $\langle g, \zeta \rangle = \int_{\mathbb{R}_+} g(x)\,\zeta(dx)$. Our equations will involve expressions of the form $\int_{[a,\infty)} g(x-a)\,\zeta(dx)$, for $a \geq 0$. To ease notation throughout, we write this as $\langle g(\cdot - a), \zeta\rangle$, adopting the convention that g is always extended to be identically zero on $(-\infty, 0)$. The set $\mathcal{M}_F$ is endowed with the topology associated with weak convergence of measures. Recall that, for $\{\zeta_n,\ n \in \mathbb{N}\} \subset \mathcal{M}_F$ and $\zeta \in \mathcal{M}_F$, ζ_n converges to ζ weakly as $n \to \infty$ if and only if $\langle g, \zeta_n\rangle \to \langle g, \zeta\rangle$ as $n \to \infty$ for all $g \in \mathbf{C}_b(\mathbb{R}_+)$, in which case we write $\zeta_n \xrightarrow{w} \zeta$ as $n \to \infty$. With this topology, $\mathcal{M}_F$ is a Polish space (cf. Prohorov [14]). A family $\{\zeta_t,\ t \in (0,\infty)\} \subset \mathcal{M}_F$ converges to $\zeta \in \mathcal{M}_F$ if and only if for each sequence $\{t_n,\ n \in \mathbb{N}\} \subset (0,\infty)$ such that $t_n \to \infty$ as $n \to \infty$ we have $\zeta_{t_n} \xrightarrow{w} \zeta$ as $n \to \infty$. We denote the zero measure in $\mathcal{M}_F$ by $\mathbf{0}$ and the measure in $\mathcal{M}_F$ that puts one unit of mass at the point $x \in \mathbb{R}_+$ by δ_x.

We will use "$\Rightarrow$" to denote convergence in distribution of random elements of a metric space. Following Billingsley [1], we will use $\mathbf{P}$ and $\mathbf{E}$ respectively to denote the probability measure and expectation operator associated with whatever space the relevant random element is defined on. All stochastic processes used in this paper will be assumed to have paths that are right continuous with finite left limits (r.c.l.l.). For a Polish space $\mathcal{S}$, we denote by $D([0,\infty), \mathcal{S})$ the space of r.c.l.l. functions from $[0,\infty)$ into $\mathcal{S}$, and we endow this space with the usual Skorohod J_1-topology (cf. Ethier and Kurtz [5]). The subspace of continuous functions from $[0,\infty)$ into $\mathcal{S}$ will be denoted by $C([0,\infty), \mathcal{S})$. When restricted to this space, the Skorohod J_1-topology is the same as that induced by uniform convergence on compact time intervals.

2.2. The processor sharing queue. Here, we describe our processor sharing queueing system. The primitive stochastic processes and initial condition for our model are introduced in §2.2.1. The system dynamics and performance processes are described in §2.2.2. Here, an important quantity for the processor sharing queue is introduced, namely the cumulative service process. In §2.2.3, we introduce the state descriptor and a dynamic equation associated with its evolution. The state descriptor and the associated dynamic equation play a fundamental role in justifying the strictly supercritical fluid model as a first-order approximation of an overloaded processor sharing queue.

2.2.1. Primitive processes and initial condition. The *exogenous arrival process* $E(\cdot)$ is a rate α delayed-renewal process. The arrival rate α is assumed to be strictly positive and finite. Jump times for $E(\cdot)$ correspond to times at which jobs enter the system. This renewal process is defined from a sequence of *interarrival times* $\{u_i\}_{i=1}^\infty$, where u_1 denotes the time at which the first job to arrive after time zero enters the system and u_i, $i \geq 2$, denotes the time between the arrival of the $(i-1)$st and ith jobs to enter the system after time zero. Frequently, we will simply refer to the ith job to enter the system after time zero as the ith arrival. Thus, $U_i = \sum_{j=1}^i u_j$ is the time at which the ith arrival enters the system, which is interpreted as zero if $i = 0$, and $E(t) = \sup\{i \geq 0\colon U_i \leq t\}$ is the number of exogenous arrivals by time t. We assume that the sequence $\{u_i\}_{i=2}^\infty$ is an i.i.d. sequence of nonnegative random variables with $\mathbf{E}[u_2] = 1/\alpha$. The random variable u_1 is associated with an initial delay preceding the first arrival and is assumed to be a strictly positive random variable, independent of $\{u_i\}_{i=2}^\infty$, with finite mean, but otherwise may have an arbitrary distribution. We refer to u_1 as the *initial residual interarrival time*.

The *service process*, $\{V(i)\}_{i=0}^\infty$, is such that $V(i)$ records the total amount of time required from the server to process the first i arrivals. More precisely, let $\{v_i\}_{i=1}^\infty$ denote an i.i.d. sequence of strictly positive random variables with common distribution given by a Borel probability measure ν on $\mathbb{R}_+$. We interpret v_i as the amount of time required from the server to process the ith arrival. The v_is are known as the *service times* and ν is known as the *service time distribution*. Then, $V(i) = \sum_{j=1}^i v_j$, $i \geq 0$, which is taken to be zero if $i = 0$. It is assumed that $v_1 > 0$ and $\mathbf{E}[v_1] < \infty$. In terms of ν, these assumptions are expressed by saying that ν does not charge the origin ($\nu(\{0\}) = 0$) and $\langle \chi, \nu \rangle < \infty$. (Recall that $\chi(x) = x$ for all $x \in \mathbb{R}_+$.)

The two processes $E(\cdot)$ and $V(\cdot)$ are called the *primitive processes*, since they provide the primitive stochastic inputs for the model. The processes $E(\cdot)$ and $V(\cdot)$ are not assumed to be independent of one another since our interest is in asymptotic behavior under fluid scaling where only laws of large numbers come into play.

Any job that is present in the system at time zero is called an *initial job*. The *initial condition* specifies the number $Z(0)$ of initial jobs and each initial job's respective service requirement. Here, $Z(0)$ is assumed to be a nonnegative, integer-valued random variable with finite mean. The service times for the initial jobs are taken to be the first $Z(0)$ elements of a sequence $\{\tilde{v}_j\}_{j=1}^{\infty}$ of strictly positive random variables, with $\tilde{v}_j$ being the service time requirement of the jth initial job, $1 \le j \le Z(0)$. It is assumed that the initial workload has a finite mean, i.e., that $\mathbf{E}\big[\sum_{j=1}^{Z(0)} \tilde{v}_j\big] < \infty$. The random variables $Z(0)$ and $\{\tilde{v}_j\}_{j=1}^{\infty}$ are neither assumed to be independent of one another nor of the primitive processes.

2.2.2. Performance processes and descriptive equations. In a processor sharing queue, the server, rather than providing service to just one job at a time, works simultaneously on all jobs currently in the system by providing an equal fraction of its attention to each. In particular, at any given time that the system is nonempty, each job in the system is being processed at a rate that is the reciprocal of the number of jobs in the system. When the server has fulfilled a given job's service requirement, the job exits the system. This system is known as a processor sharing queue.

As the processor sharing queue evolves in time, certain r.c.l.l. stochastic processes are used to track important measures of performance for the system, e.g., queue length, workload, and idle time. Let $Z(t)$ denote the queue length at time t, which is the total number of jobs in the system at time t. Also, let $W(t)$ denote the (immediate) workload at time t, which is the total amount of time that the server must work in order to satisfy the remaining service requirement of each job present in the system at time t, ignoring future arrivals. Finally, let $Y(t)$ denote the cumulative amount of time that the server has been idle up to time t. The processes $W(\cdot)$, $Y(\cdot)$, and $Z(\cdot)$ are called *performance processes*. These processes satisfy a set of descriptive equations, which we now present.

We begin with the familiar equations for the *workload* $W(\cdot)$ and *idle time* $Y(\cdot)$ processes, which are valid for any nonidling service discipline, including processor sharing. For $t \ge 0$,

$$W(0) = \sum_{j=1}^{Z(0)} \tilde{v}_j, \tag{1}$$

$$W(t) = W(0) + V(E(t)) - t + Y(t), \tag{2}$$

$$Y(t) = \sup\{(W(0) + V(E(s)) - s)^-\colon 0 \le s < t\}. \tag{3}$$

A set of equations satisfied by the *queue length* process $Z(\cdot)$ under a processor sharing service discipline is the following. For $t \ge 0$,

$$Z(t) = Z(0) + E(t) - D(t), \tag{4}$$

$$D(t) = \sum_{j=1}^{Z(0)} 1_{\{\tilde{v}_j \le S(t)\}} + \sum_{i=1}^{E(t)} 1_{\{v_i \le S(t) - S(U_i)\}}, \tag{5}$$

$$S(t) = \int_0^t \varphi(Z(s))\, ds. \tag{6}$$

Recall that $\varphi(x) = 1/x$ if $x > 0$, and $\varphi(0) = 0$. The process $D(\cdot)$ is the *departure process*, where $D(t)$ represents the total number of jobs that have departed from the system by time t. The process $S(\cdot)$ is known as the *cumulative service process*, and $S(t)$ represents the cumulative amount of service time allocated per job up to time t.

The cumulative service process $S(\cdot)$ will play a particularly important role in our analysis. We will find it convenient to have notation for the increments of this process. For $t, h \ge 0$, define the *cumulative service per job in* $[t, t+h]$ by

$$S_{t,\, t+h} = S(t+h) - S(t) = \int_t^{t+h} \varphi(Z(s))\, ds. \tag{7}$$

For $t \ge U_i$, the ith arrival receives an amount of service equal to $v_i \wedge S_{U_i,\, t}$ by time t. Define the *residual service time* at time $t \ge 0$ of the ith arrival, $i \in \{1, \ldots, E(t)\}$, and of the jth initial job, $j \in \{1, \ldots, Z(0)\}$, by

$$R_i(t) = (v_i - S_{U_i,\, t})^+ \qquad \text{and} \qquad \tilde{R}_j(t) = (\tilde{v}_j - S(t))^+, \tag{8}$$

respectively. The workload at time $t \geq 0$ can be rewritten as

$$W(t) = \sum_{j=1}^{Z(0)} \tilde{R}_j(t) + \sum_{i=1}^{E(t)} R_i(t). \tag{9}$$

2.2.3. Measure-valued state descriptor. As in Gromoll et al. [9], we use a measure-valued process to keep track of the residual service times. For each $t \geq 0$, let $\mu(t)$ be the random, finite, Borel measure on $\mathbb{R}_+ = [0, \infty)$ given by

$$\mu(t) = \sum_{j=1}^{Z(0)} 1_{(0,\infty)}(\tilde{R}_j(t))\delta_{\tilde{R}_j(t)} + \sum_{i=1}^{E(t)} 1_{(0,\infty)}(R_i(t))\delta_{R_i(t)}. \tag{10}$$

Recall that δ_x is the measure that puts a single unit of mass at x for $x \in \mathbb{R}_+$. Thus, the random measure $\mu(t)$ has a unit of mass at the residual service time of each job that is still in the system at time t. The indicator functions in the above definition serve to eliminate jobs with zero residual service times from the description of the system state, since such jobs have departed the system. We call $\mu(t)$ the *measure-valued state descriptor* at time t. Note that the queue length and workload at time t can be obtained from the state descriptor by integrating against an appropriate function. In particular, for $t \geq 0$,

$$Z(t) = \langle 1, \mu(t) \rangle \qquad \text{and} \qquad W(t) = \langle \chi, \mu(t) \rangle. \tag{11}$$

Furthermore, given the primitive processes and the initial condition, one can recover $D(\cdot)$ and $S(\cdot)$ from $Z(\cdot)$, and $Y(\cdot)$ from $W(\cdot)$. Notice that the information given by the initial condition is described by the random initial measure $\mu(0)$. In particular, our assumptions on the initial condition are given by

$$\mathbf{E}[Z(0)] = \mathbf{E}[\langle 1, \mu(0) \rangle] < \infty \qquad \text{and} \qquad \mathbf{E}[W(0)] = \mathbf{E}[\langle \chi, \mu(0) \rangle] < \infty. \tag{12}$$

Henceforth, these will be stated in terms of $\mu(0)$, rather than in terms of $Z(0)$ and $\{\tilde{v}_j\}_{j=1}^{\infty}$.

The assumptions on the initial condition and primitive processes, together with the processor sharing dynamics, imply that, for each $t \geq 0$, the random measure $\mu(t)$ is an element of $\mathcal{M}_F$, the space of finite, nonnegative Borel measures on $\mathbb{R}_+$. It is straightforward to see that $\mu(\cdot)$ is a measure-valued stochastic process with sample paths in the Polish space $D([0, \infty), \mathcal{M}_F)$ of functions from $[0, \infty)$ into $\mathcal{M}_F$ that are right continuous with finite left limits, endowed with the Skorohod J_1-topology (cf. Ethier and Kurtz [5]). This measure-valued process and its fluid limit are the objects of central interest in this paper. An equivalent formulation of (10) is to consider the real-valued processes $\langle g, \mu(\cdot) \rangle$ for a suitable class of functions g: $\mathbb{R}_+ \to \mathbb{R}$. In fact, (10) holds if and only if for each bounded, Borel-measurable function g: $\mathbb{R}_+ \to \mathbb{R}$, the process $\mu(\cdot)$ satisfies

$$\langle g, \mu(t) \rangle = \sum_{j=1}^{Z(0)} (1_{(0,\infty)} g)(\tilde{R}_j(t)) + \sum_{i=1}^{E(t)} (1_{(0,\infty)} g)(R_i(t)), \quad \text{for all } t \geq 0. \tag{13}$$

The set of equations given by (13), or equivalently equation (10), will be the starting point for our analysis of processor sharing queues.

3. Supercritical fluid model and main results.

3.1. Definition of fluid model solutions. The fluid model has two parameters, $\alpha \in (0, \infty)$ and a Borel probability measure ν on $\mathbb{R}_+$ that does not charge the origin and satisfies $\langle \chi, \nu \rangle < \infty$. These parameters are limits of parameters in the queueing system, where α corresponds to the rate at which jobs arrive to the system and the probability measure ν corresponds to the distribution of the i.i.d. service times for those jobs. The *traffic intensity parameter* ρ is given by $\rho = \alpha/\beta$, where $\beta = 1/\langle \chi, \nu \rangle$. The pair (α, ν) is referred to as the fluid model *data*, or simply the data. Here we only consider *supercritical data*, i.e., data that satisfies $\rho \geq 1$. The adjectives *critical* and *strictly supercritical* are used to refer to data that satisfies $\rho = 1$ and $\rho > 1$, respectively.

The fluid model is formulated by considering a formal law of large numbers limit $\bar{\mu}(\cdot)$ for the measure-valued processes corresponding to a sequence of processor sharing queues that satisfies appropriate asymptotic conditions. The dynamics of $\bar{\mu}(\cdot)$ are prescribed through a set of so-called *fluid model equations* that are satisfied by the real-valued projections $\langle g, \bar{\mu}(t) \rangle$, $t \geq 0$, for each function g: $\mathbb{R}_+ \to \mathbb{R}$ in a suitable class. To avoid the singular behavior associated with the abrupt departure of mass at the origin, this class is chosen so that the functions, together with their first derivatives, vanish at the origin. Specifically, we work with the class

$$\mathscr{C} = \{g \in \mathbf{C}_b^1(\mathbb{R}_+)\colon g(0) = 0,\ g'(0) = 0\}, \tag{14}$$

which is large enough for the purpose of characterizing fluid model solutions. The form of the fluid model equations depends on the data (α, ν), or, more precisely, on the value of the traffic intensity parameter ρ.

We begin by recalling the formulation of the critical fluid model used in Gromoll et al. [9].

Definition 3.1. Given critical data (α, ν) (i.e., data satisfying $\rho = 1$), a fluid model solution for the data (α, ν) is a function $\bar{\mu}\colon [0, \infty) \to \mathscr{M}_F$ such that

(C.1) $\bar{\mu}(\cdot)$ is continuous,
(C.2) $\langle 1_{\{0\}}, \bar{\mu}(t)\rangle = 0$ for all $t \geq 0$,
(C.3) for all $g \in \mathscr{C}$, $\bar{\mu}(\cdot)$ satisfies

$$\langle g, \bar{\mu}(t)\rangle = \langle g, \bar{\mu}(0)\rangle - \int_0^t \frac{\langle g', \bar{\mu}(s)\rangle}{\langle 1, \bar{\mu}(s)\rangle}\, ds + \alpha t \langle g, \nu\rangle, \tag{15}$$

for all $t < t^* = \inf\{s \geq 0\colon \bar{\mu}(s) = \mathbf{0}\}$, and
(C.4) for all $t \geq t^*$, $\bar{\mu}(t) = \mathbf{0}$.

The equations in (15), one for each g, are called the *critical fluid model equations*. The reader is referred to Gromoll et al. [9] for an informal interpretation of the conditions in this definition. Of particular relevance here is the interpretation of (C.4), which hinges on the condition $\rho = 1$. Specifically, when $\rho = 1$, the service rate is the same as the arrival rate. Therefore, in the fluid model, work should not build up from the zero initial measure, which is guaranteed by (C.4). Moreover, Gromoll et al. [9, Lemma 4.4] implies that if the initial measure is nonzero, then $t^* = \infty$, and consequently, (15) holds for all $t \geq 0$. Thus, when $\rho = 1$, (C.4) does not take effect for fluid model solutions with nonzero initial measures.

Next, we give the formulation of the strictly supercritical fluid model. In this case, the arrival rate exceeds the service rate. Therefore, one does expect mass to build up from the zero initial measure. In particular, (C.4) does not apply, and the above definition must be modified.

Definition 3.2. Given strictly supercritical data (α, ν) (i.e., data satisfying $\rho > 1$), a fluid model solution for the data (α, ν) is a function $\bar{\mu}\colon [0, \infty) \to \mathscr{M}_F$ such that

(S.1) $\bar{\mu}(\cdot)$ is continuous,
(S.2) for all $t \geq 0$, $\langle 1_{\{0\}}, \bar{\mu}(t)\rangle = 0$, and
(S.3) for all $g \in \mathscr{C}$, $\bar{\mu}(\cdot)$ satisfies

$$\langle g, \bar{\mu}(t)\rangle = \langle g, \bar{\mu}(0)\rangle - \int_0^t \langle g', \bar{\mu}(s)\rangle \varphi(\langle 1, \bar{\mu}(s)\rangle)\, ds + \alpha t\langle g, \nu\rangle, \quad \text{for all } t \geq 0. \tag{16}$$

Recall that $\varphi(x) = 1/x$ if $x \in (0, \infty)$ and $\varphi(0) = 0$. The equations in (16), one for each g, are called the *strictly supercritical fluid model equations*. The function $\varphi(\cdot)$ is used in (16) to allow for the possibility that $\bar{\mu}(s) = \mathbf{0}$ for some $s \geq 0$. Note that the integral term in (16) is well defined since $\bar{\mu}(\cdot)$ is continuous and, for each $g \in \mathscr{C}$, the absolute value of the integrand is bounded above by $\|g'\|_\infty$.

Given supercritical data (α, ν) (i.e., data satisfying $\rho \geq 1$), the terminology *a fluid model solution for the data* (α, ν) will mean a function $\bar{\mu}(\cdot)$ satisfying Definition 3.1 if $\rho = 1$ or Definition 3.2 if $\rho > 1$. We refer to a function $\bar{\mu}(\cdot)$ as *a fluid model solution* if there exists supercritical data (α, ν) such that $\bar{\mu}(\cdot)$ is a fluid model solution for the data (α, ν).

Let $\bar{\mu}(\cdot)$ be a fluid model solution. The fluid analogue of the queue length process is given by

$$\bar{Z}(t) = \langle 1, \bar{\mu}(t)\rangle, \quad \text{for all } t \geq 0. \tag{17}$$

For obvious reasons, $\bar{Z}(t)$ is referred to as the *total mass* at time t. Due to the assumed continuity of fluid model solutions, $\bar{Z}(\cdot)$ is continuous. The fluid analogue of the cumulative service per job is defined by

$$\bar{S}(t) = \int_0^t \varphi(\bar{Z}(u))\, du, \quad \text{for all } t \geq 0, \tag{18}$$

where $\bar{S}(t) = \infty$ if the integral on the right diverges. The fluid analogue of the workload process is given by

$$\bar{W}(t) = \langle \chi, \bar{\mu}(t)\rangle, \quad \text{for all } t \geq 0. \tag{19}$$

If, for a particular $t \geq 0$, we have $\langle \chi, \bar{\mu}(t)\rangle = \infty$, then $\bar{W}(t) = \infty$. Also let

$$\bar{M}(t, x) = \langle 1_{(x,\infty)}, \bar{\mu}(t)\rangle, \quad \text{for all } t \geq 0 \text{ and } x \in \mathbb{R}_+. \tag{20}$$

Then, it is easily verified that $\|\bar{M}(t,\cdot)\|_{L^1} = \bar{W}(t)$ for all $t \geq 0$.

We note that, if $\rho > 1$, then it follows from (16) that

$$\bar{Z}(t) > 0, \quad \text{for all } t > 0. \tag{21}$$

This can be seen by substituting a sequence of nonnegative functions $\{g_n\}_{n\in\mathbb{N}} \subset \mathscr{C}$ into (16), where $g_n \nearrow \chi$ and $g_n' \nearrow 1_{(0,\infty)}$ as $n \to \infty$. Using monotone convergence, it follows that

$$\bar{W}(t) = \bar{W}(0) - \int_0^t \bar{Z}(s)\varphi(\bar{Z}(s))\, ds + \rho t, \quad \text{for all } t \geq 0, \tag{22}$$

which implies that $\bar{W}(t) \geq \bar{W}(0) - t + \rho t > 0$ for all $t > 0$ and verifies (21). Further, note that (21) together with (22) implies that if $\rho > 1$, then

$$\bar{W}(t) = \bar{W}(0) + (\rho - 1)t, \quad \text{for all } t \geq 0, \tag{23}$$

which is expected since this should hold for all nonidling disciplines. Finally, if $\rho > 1$ and $\bar{Z}(0) > 0$, then $\bar{Z}(t)$ is continuous and strictly positive for all $t \geq 0$ (cf. (21)), and $\bar{S}(\cdot) \in \mathbf{C}^1([0,\infty))$ with

$$\frac{d}{dt}\bar{S}(t) = \frac{1}{\bar{Z}(t)}, \quad \text{for all } t \geq 0.$$

3.2. Properties of fluid model solutions.

3.2.1. No atoms property. It will be shown that fluid model solutions have no atoms for all time. In particular, fluid model solutions take values in the following set:

$$\mathscr{M}_F^c = \{\xi \in \mathscr{M}_F \colon \xi(\{x\}) = 0 \text{ for all } x \in \mathbb{R}_+\},$$

and in

$$\mathscr{M}_F^{c,p} = \{\xi \in \mathscr{M}_F^c \colon \xi \neq \mathbf{0}\},$$

when the initial measure is nonzero. Intuitively, this is so because fluid model solutions are continuous (cf. (C.1) and (S.1)). Indeed, given $x \in (0,\infty)$ and $t \geq 0$, one expects that all of the mass located at x at time t will depart the system simultaneously. But then, if a fluid model solution were to have an atom at x at time t, the total mass would have a downward jump (i.e., a discontinuity) at the departure time of that atom, which would contradict the assumption that fluid model solutions are continuous. This intuition is made rigorous for strictly supercritical data, nonzero initial measures, and $t = 0$ in the proof of Proposition 4.1 below. For strictly supercritical (resp. critical) data and nonzero initial measures, the no atoms property follows from Corollary 4.1 (resp. Remark 4.2). For the zero initial measure, the no atoms property is immediate from (C.4) for critical data and from Theorem 3.2 and the fact that the measure ς (defined in (29) below) has no atoms for strictly supercritical data.

3.2.2. Existence and uniqueness theorem for strictly supercritical data.

THEOREM 3.1. *Given strictly supercritical data* (α, ν) *and* $\xi \in \mathscr{M}_F^c$, *a fluid model solution* $\bar{\mu}(\cdot)$ *for the data* (α,ν) *such that* $\bar{\mu}(0) = \xi$ *exists and is unique. Moreover,*

$$\langle \chi, \bar{\mu}(t)\rangle = (\rho - 1)t + \langle \chi, \xi\rangle, \quad \textit{for all } t \geq 0. \tag{24}$$

Equation (24) is called the *fluid workload equation*. If $\langle \chi, \xi\rangle = \infty$, then (24) is interpreted as saying that $\langle \chi, \bar{\mu}(t)\rangle = \infty$ for all $t \geq 0$. Equation (24) was already verified in (23), but the proof of existence and uniqueness still remains. Under the assumption that the data is critical, the analogue of Theorem 3.1 was proved in Gromoll et al. [9, §4] (cf. [9, Theorem 3.1]). To prove Theorem 3.1, there are two cases to consider; $\xi \in \mathscr{M}_F^{c,p}$ and $\xi = \mathbf{0}$. For $\xi \in \mathscr{M}_F^{c,p}$, since any fluid model solution $\bar{\mu}(\cdot)$ for strictly supercritical data (α,ν) such that $\bar{\mu}(0) = \xi$ is continuous and never hits the zero measure (cf. (21)), it follows that $t^* \equiv \inf\{t \geq 0 \colon \bar{\mu}(t) = \mathbf{0}\} = \infty$ and that $\bar{\mu}(\cdot)$ satisfies (15), i.e., $\bar{\mu}(\cdot)$ satisfies the same conditions as for the critical fluid model. In §4.1, we will see that because of this, the proof of existence and uniqueness given in Gromoll et al. [9] for critical data extends to strictly supercritical data for $\bar{\mu}(0) \neq \mathbf{0}$ (cf. Theorem 4.1 below). Thus, the main contribution here is the proof of Theorem 3.1 for $\xi = \mathbf{0}$, which follows as a consequence of Theorem 3.2 below. Theorems 3.2 and 4.1 imply Theorem 3.1.

3.2.3. Starting from the zero initial measure for strictly supercritical data. If the data (α, ν) is strictly supercritical, then we expect mass in the fluid model to exit from the system at a rate that is strictly slower than the rate at which it arrives to the system. Therefore, the total mass is expected to grow. The next theorem, Theorem 3.2 below, validates this intuition, and moreover, explicitly describes the distribution of the mass as it builds up from the zero measure.

For the statement of Theorem 3.2, we will need to provide some background. Note that, given strictly supercritical data (α, ν), there is a unique, positive real number m such that

$$\alpha(1 - \langle \exp(-m\,\cdot), \nu\rangle) = m. \tag{25}$$

To verify this, let ν_e denote the *excess lifetime probability measure* associated with ν. Specifically, ν_e is the Borel probability measure on $\mathbb{R}_+$ that is absolutely continuous with respect to Lebesgue measure and has density function

$$f_e(x) = \beta\langle 1_{(x,\infty)}, \nu\rangle, \quad \text{for all } x \in \mathbb{R}_+. \tag{26}$$

Recall that $\beta = 1/\langle \chi, \nu\rangle$. The left side of (25) can be expressed in terms of ν_e since

$$\alpha(1 - \langle \exp(-y\cdot), \nu\rangle) = \alpha\langle 1 - \exp(-y\cdot), \nu\rangle = \rho y\langle \exp(-y\cdot), \nu_e\rangle, \quad \text{for all } y \in \mathbb{R}_+.$$

Here the first equality makes use of the fact that ν is a probability measure and the second equality follows by substituting $y\int_0^{\cdot} \exp(-yv)\,dv$ for $1 - \exp(-y\cdot)$ and then interchanging the order of integration. So we see that, for $m \in (0, \infty)$, (25) holds if and only if

$$\langle \exp(-m\cdot), \nu_e\rangle = \rho^{-1}. \tag{27}$$

There is a unique $m \in (0, \infty)$ satisfying (27) since $\rho^{-1} \in (0, 1)$ and since, as a function of $m \in (0, \infty)$, the left member of (27) is strictly decreasing and maps onto $(0, 1)$. The real number m can be thought of as a type of Malthusian parameter (Smith [17]).

Given strictly supercritical data (α, ν), let m be the unique, positive real number satisfying (25) and define $p\colon \mathbb{R}_+ \to \mathbb{R}_+$ by

$$p(x) = \alpha\langle 1_{(x,\infty)}(\cdot) \exp(m(x - \cdot)), \nu\rangle, \quad \text{for all } x \in \mathbb{R}_+. \tag{28}$$

Note that $\int_0^\infty p(x)\,dx = 1$. To see this, integrate (28) over $\mathbb{R}_+$, interchange the order of integration, simplify, and then use (25). Let $\varsigma \in \mathcal{M}_F$ denote the measure that is absolutely continuous with respect to Lebesgue measure and has Radon-Nikodym derivative $mp(\cdot)$:

$$\varsigma(dx) = mp(x)\,dx, \quad \text{for all } x \in \mathbb{R}_+. \tag{29}$$

In particular,

$$\langle 1, \varsigma\rangle = m. \tag{30}$$

Finally, let $\bar{\varsigma}\colon [0, \infty) \to \mathcal{M}_F$ be given by

$$\bar{\varsigma}(t) = t\varsigma, \quad \text{for all } t \geq 0. \tag{31}$$

THEOREM 3.2. *Given strictly supercritical data (α, ν), let $\bar{\varsigma}(\cdot)$ be defined via (28), (29), and (31). Then, $\bar{\varsigma}(\cdot)$ is the unique fluid model solution for the data (α, ν) with initial measure* **0**.

As previously noted, Theorem 3.2 explicitly describes the distribution of mass as it builds up from zero. In particular, the distribution is invariant in shape (where the shape has probability density function p) and the total mass grows linearly in t. Lemma 4.4 in §4.2 implies that $\bar{\varsigma}(\cdot)$ is a fluid model solution for the data (α, ν). The more challenging issue is to verify the uniqueness claimed in Theorem 3.2; this is done in §4.4.3.

3.2.4. Continuous dependence on the initial measure. Given supercritical data (α, ν) and $\xi \in \mathcal{M}_F^c$, let $\bar{\mu}_\xi(\cdot)$ denote the unique fluid model solution for data (α, ν) with initial measure ξ. Let $\Xi\colon \mathcal{M}_F^c \to C([0, \infty), \mathcal{M}_F)$ be such that

$$\Xi(\xi) = \bar{\mu}_\xi(\cdot).$$

REMARK 3.1. Since $C([0, \infty), \mathcal{M}_F)$ is a subspace of $D([0, \infty), \mathcal{M}_F)$, one could use the latter in the above definition of Ξ, as was done in Gromoll et al. [9].

Theorem 3.3. *Let (α, ν) be supercritical data. Then, Ξ is continuous.*

Under the condition that (α, ν) is critical data, Gromoll et al. [9, Lemma 4.9] asserts that the restriction of Ξ to $\mathcal{M}_F^{c,p}$ is continuous. As we will see in §4.1, the proof of [9, Lemma 4.9] extends to strictly supercritical data as well (cf. Lemma 4.3 below). Therefore, the main contribution here is to prove the following theorem which implies that, for all supercritical data, Ξ is continuous at the zero measure.

Theorem 3.4. *Let (α, ν) be supercritical data, and let $\{\xi_n, n \in \mathbb{N}\} \subset \mathcal{M}_F^{c,p}$ be such that $\xi_n \xrightarrow{w} \mathbf{0}$ as $n \to \infty$. Then, $\Xi(\xi_n)$ converges to $\Xi(\mathbf{0})$ in $C([0,\infty), \mathcal{M}_F)$ as $n \to \infty$.*

Theorem 3.4 is proved in §4.5.2. Together Gromoll et al. [9, Lemma 4.9] (for critical data), Lemma 4.3 (for strictly supercritical data), and Theorem 3.4 (for supercritical data) imply Theorem 3.3.

3.2.5. Shift and scaling properties of fluid model solutions. Let (α, ν) be supercritical data and $\bar{\mu}(\cdot)$ be a fluid model solution for the data (α, ν). For $s \geq 0$, define $\tau_s \bar{\mu}\colon [0,\infty) \to \mathcal{M}_F$ to be given by

$$(\tau_s \bar{\mu})(t) = \bar{\mu}(t+s) \quad \text{for all } t \geq 0. \tag{32}$$

For $s > 0$, define $\mathcal{L}_s \bar{\mu}\colon [0,\infty) \to \mathcal{M}_F$ to be such that

$$(\mathcal{L}_s \bar{\mu})(t) = \frac{\bar{\mu}(st)}{s} \quad \text{for all } t \geq 0. \tag{33}$$

The following lemma is straightforward to verify.

Lemma 3.1. *Let (α, ν) be supercritical data, and $\bar{\mu}(\cdot)$ be a fluid model solution for the data (α, ν).*
(i) *For each $s \geq 0$, $(\tau_s \bar{\mu})(\cdot)$ is a fluid model solution for the data (α, ν).*
(ii) *For each $s > 0$, $(\mathcal{L}_s \bar{\mu})(\cdot)$ is a fluid model solution for the data (α, ν).*

We refer to Lemma 3.1(i) as the *shift property* and to Lemma 3.1(ii) as the *scaling property*. When combined with continuity in the initial condition, the shift and scaling properties have significant consequences for the behavior of fluid model solutions. These are described in §§3.2.6 and 3.2.7 below.

3.2.6. Asymptotics of fluid model solutions for strictly supercritical data.

Theorem 3.5. *Let (α, ν) be strictly supercritical data, and let ς be defined by (28) and (29). Any fluid model solution $\bar{\mu}(\cdot)$ for the data (α, ν) satisfies*

$$\frac{\bar{\mu}(t)}{t} \xrightarrow{w} \varsigma, \quad \text{as } t \to \infty. \tag{34}$$

Proof. Let (α, ν) be strictly supercritical data and $\bar{\mu}(\cdot)$ be a fluid model solution for the data (α, ν). By Lemma 3.1(ii), for each $s > 0$, $(\mathcal{L}_s \bar{\mu})(\cdot)$ is a fluid model solution for the data (α, ν). Note that, for each $s > 0$, $\bar{\mu}(s)/s = (\mathcal{L}_s \bar{\mu})(1)$ (cf. (33)). Thus, it suffices to show that

$$(\mathcal{L}_s \bar{\mu})(1) \xrightarrow{w} \varsigma, \quad \text{as } s \to \infty. \tag{35}$$

Also note that, for each $s > 0$, $(\mathcal{L}_s \bar{\mu})(0) = \bar{\mu}(s)/s$ and so,

$$(\mathcal{L}_s \bar{\mu})(0) \xrightarrow{w} \mathbf{0}, \quad \text{as } s \to \infty.$$

If $\bar{\mu}(0) = \mathbf{0}$, then, by Theorem 3.2 and (31), $(\mathcal{L}_s \bar{\mu})(0) \in \mathcal{M}_F^{c,p}$ for each $s > 0$. Similarly, if $\bar{\mu}(0) \neq \mathbf{0}$, then by Proposition 4.1 below, $(\mathcal{L}_s \bar{\mu})(0) \in \mathcal{M}_F^{c,p}$ for each $s > 0$. Thus, from Theorem 3.4, it follows that

$$(\mathcal{L}_s \bar{\mu})(1) = \Xi((\mathcal{L}_s \bar{\mu})(0))(1) \xrightarrow{w} \Xi(\mathbf{0})(1) = \bar{\varsigma}(1), \quad \text{as } s \to \infty.$$

Since $\bar{\varsigma}(1) = \varsigma$ (cf. (31)), (35) follows. $\square$

As a consequence of Theorem 3.5 and (30), it follows that if $\bar{\mu}(\cdot)$ is a fluid model solution for the data (α, ν), then

$$\lim_{t\to\infty} \frac{\bar{Z}(t)}{t} = m.$$

Under the condition that $\overline{W}(0) < \infty$, this was previously recognized in Chen et al. [2, Proposition 6]. Note that the asymptotic linear rate of growth for the fluid analogue of the queue length in a first-in-first-out queue is $\alpha - \beta$. If ν is exponential, then $m = \alpha - \beta$, but $m \neq \alpha - \beta$ in general.

A statement analogous to that in Theorem 3.5 can be made when the data is critical (cf. Puha and Williams [15, Theorem 1.2]). In particular, for $\xi \in \mathcal{M}_F^c$ such that $\langle \chi, \xi\rangle < \infty$, $\bar{\mu}(t) \xrightarrow{w} \beta_e \langle \chi, \xi\rangle \nu_e$ as $t \to \infty$, with $\beta_e = 1/\langle \chi, \nu_e\rangle$ if $\langle \chi, \nu_e\rangle < \infty$ and $\beta_e = 0$ otherwise. For this, there is no need to normalize by a linear growth factor since the total mass does not grow with time. In the critical case, the probability density function $f_e(\cdot)$ plays the role that the probability density function $p(\cdot)$ plays in the strictly supercritical case. Note that if ρ decreases to one by letting α decrease to β, then m decreases to zero and $p(\cdot)$ given by (28) converges to $f_e(\cdot)$. In this way, $p(\cdot)$ can be viewed as a generalization of $f_e(\cdot)$ suited to the strictly supercritical setting.

3.2.7. Stationary fluid model solutions for strictly supercritical data. For any fluid model solution, Theorem 3.5 implies that for strictly supercritical data the fluid analogue of the queue length process is asymptotically linear, and that the fluid analogue of the process giving the empirical distribution of the residual service times converges to $m^{-1}\varsigma$ as time tends to infinity. This naturally raises the question as to what conditions are necessary and sufficient for this last fluid analogue to be identically equal to $m^{-1}\varsigma$, and motivates the following definition.

Definition 3.3. Let (α, ν) be strictly supercritical data. A fluid model solution $\bar{\mu}(\cdot)$ for the data (α, ν) is stationary (in shape) if there exists a Borel probability measure π on $\mathbb{R}_+$ such that $\bar{\mu}(t) = \langle 1, \bar{\mu}(t)\rangle \pi$ for all $t \geq 0$.

Given strictly supercritical data (α, ν), $\bar{\varsigma}(\cdot)$ defined by (28), (29), and (31) is an example of a stationary fluid model solution for the data (α, ν). From $\bar{\varsigma}(\cdot)$, it is easy to construct a one-parameter family of stationary fluid model solutions for the data (α, ν). Specifically, by Lemma 3.1(i), for each $s \geq 0$, $(\tau_s \bar{\varsigma})(\cdot)$ is a fluid model solution for the data (α, ν), which is easily verified to be stationary. The next theorem implies that this one-parameter family contains all of the stationary fluid model solutions for the data (α, ν). In other words, the probability density function $p(\cdot)$ defined by (28) determines the unique stationary shape for the data (α, ν).

Theorem 3.6. *Let (α, ν) be strictly supercritical data, and let $\bar{\varsigma}(\cdot)$ be defined by (28), (29), and (31). A fluid model solution $\bar{\mu}(\cdot)$ for the data (α, ν) is stationary if and only if for some $s \in [0, \infty)$, $\bar{\mu}(t) = (\tau_s \bar{\varsigma})(t)$ for all $t \geq 0$.*

Proof. Let (α, ν) be strictly supercritical data and let $\bar{\varsigma}(\cdot)$ be defined by (28), (29), and (31). As noted above, for any $s \in [0, \infty)$, $(\tau_s \bar{\varsigma})(\cdot)$ is a stationary fluid model solution for the data (α, ν). We must show that these are the only stationary fluid model solutions for the data (α, ν). For this, suppose that $\bar{\mu}(\cdot)$ is a stationary fluid model solution for the data (α, ν). Then, there exists a Borel probability measure π on $\mathbb{R}_+$ such that, for all $t \geq 0$,

$$\bar{\mu}(t) = \langle 1, \bar{\mu}(t)\rangle \pi. \tag{36}$$

By Theorem 3.5, $\lim_{t\to\infty} t^{-1}\bar{\mu}(t) = \varsigma$, where the convergence is weak convergence in $\mathcal{M}_F$. This implies that $\lim_{t\to\infty} t^{-1}\langle 1, \bar{\mu}(t)\rangle = \langle 1, \varsigma\rangle = m$. Thus, on dividing (36) by t and letting $t \to \infty$, we obtain $\pi = \varsigma/\langle 1, \varsigma\rangle = m^{-1}\varsigma$. Therefore,

$$\bar{\mu}(t) = \langle 1, \bar{\mu}(t)\rangle m^{-1}\varsigma \quad \text{for all } t \geq 0. \tag{37}$$

Hence, if it can be shown that

$$\langle 1, \bar{\mu}(t)\rangle = mt + \langle 1, \bar{\mu}(0)\rangle, \quad \text{for all } t \geq 0, \tag{38}$$

it will follow that $\bar{\mu}(\cdot) = (\tau_s \bar{\varsigma})(\cdot)$ for $s = m^{-1}\langle 1, \bar{\mu}(0)\rangle$. Equation (37) together with (16) and (21) implies that for each $g \in \mathcal{C}$,

$$\langle 1, \bar{\mu}(t)\rangle m^{-1}\langle g, \varsigma\rangle = \langle 1, \bar{\mu}(0)\rangle m^{-1}\langle g, \varsigma\rangle - tm^{-1}\langle g', \varsigma\rangle + \alpha t\langle g, \nu\rangle, \quad \text{for all } t \geq 0.$$

In §4.2, it is verified that for each $g \in \mathcal{C}$, $\langle g, \varsigma\rangle = -m^{-1}\langle g', \varsigma\rangle + \alpha\langle g, \nu\rangle$ (cf. (67)). Thus, for each $g \in \mathcal{C}$,

$$\langle 1, \bar{\mu}(t)\rangle m^{-1}\langle g, \varsigma\rangle = \langle 1, \bar{\mu}(0)\rangle m^{-1}\langle g, \varsigma\rangle + t\langle g, \varsigma\rangle, \quad \text{for all } t \geq 0.$$

Since there exists $g \in \mathcal{C}$ such that $\langle g, \varsigma\rangle \neq 0$, equation (38) follows. □

3.3. Convergence to fluid model solutions.

3.3.1. A sequence of overloaded processor sharing queues. In this section, we specify the assumptions under which the fluid limit result will be proved. Consider a sequence of processor sharing queueing models indexed by r, which increases to ∞ through a sequence in $(0, \infty)$. Each model in the sequence may be defined on a separate probability space. The rth model is defined as in §2.2, except that all accompanying processes and parameters have a superscript r appended to them. In particular, primitive processes associated with the rth system are E^r and V^r, which have parameters α^r and ν^r, respectively. Similarly, the performance processes associated with the rth system are $W^r(\cdot)$, $Y^r(\cdot)$, and $Z^r(\cdot)$, and the measure-valued process is $\mu^r(\cdot)$. Recall that the initial condition for each queueing model, as well as the assumptions placed on it, can be specified in terms of the initial random measure $\mu^r(0)$. In particular, it is assumed (cf. (12)) that for each r, $\mu^r(0)$ satisfies

$$\mathbf{E}[\langle 1, \mu^r(0)\rangle] < \infty \qquad \text{and} \qquad \mathbf{E}[\langle \chi, \mu^r(0)\rangle] < \infty. \tag{39}$$

The fluid limit result concerns the behavior of processor sharing queues on law of large numbers scale, or *fluid scale*. Accordingly, define the fluid-scaled processes

$$\bar{Z}^r(t) = \frac{1}{r} Z^r(rt) \tag{40}$$

$$\bar{W}^r(t) = \frac{1}{r} W^r(rt) \tag{41}$$

$$\bar{\mu}^r(t) = \frac{1}{r} \mu^r(rt) \tag{42}$$

$$\bar{E}^r(t) = \frac{1}{r} E^r(rt) \tag{43}$$

$$\bar{S}^r_{t,\,t+h} = S^r_{rt,\,r(t+h)} = \int_{rt}^{r(t+h)} \varphi(\langle 1, \mu^r(s)\rangle)\, ds = \int_t^{t+h} \varphi(\langle 1, \bar{\mu}^r(s)\rangle)\, ds, \tag{44}$$

for all $t \in [0, \infty)$, $h \ge 0$.

Let (α, ν) be strictly supercritical data. In order to obtain convergence in distribution of the fluid-scaled processes $\bar{\mu}^r(\cdot)$ to a process that is a.s. a fluid model solution for the data (α, ν), we impose the following asymptotic assumptions on the sequence of processor sharing queueing models. For the primitive processes, assume that as $r \to \infty$,

$$\alpha^r \longrightarrow \alpha, \tag{45}$$

$$\nu^r \xrightarrow{\text{w}} \nu, \tag{46}$$

$$\langle \chi, \nu^r\rangle \longrightarrow \langle \chi, \nu\rangle, \tag{47}$$

$$\mathbf{E}[u_1^r]/r \longrightarrow 0, \tag{48}$$

$$\mathbf{E}[u_2^r; u_2^r > r] \longrightarrow 0. \tag{49}$$

Recall that since the data is strictly supercritical, $\alpha\langle \chi, \nu\rangle > 1$. Thus, assumptions (45) and (47) guarantee that as $r \to \infty$, the systems become overloaded, i.e., that $\rho^r \to \rho > 1$ as $r \to \infty$. Assumption (48) implies that the initial residual interarrival time vanishes on fluid scale. We assume (49) in order to provide uniform control over the tail of the distribution of u_2^r, which is used to obtain a weak law of large numbers for a triangular array (cf. Gromoll et al. [9, Lemma A.2]).

For the fluid-scaled initial measures, we assume that for some random measure Θ taking values in $\mathcal{M}_F$, we have

$$(\bar{\mu}^r(0), \langle \chi, \bar{\mu}^r(0)\rangle) \Longrightarrow (\Theta, \langle \chi, \Theta\rangle), \quad \text{as } r \to \infty, \tag{50}$$

with Θ satisfying

$$\mathbf{E}[\langle 1, \Theta\rangle] < \infty \qquad \text{and} \qquad \mathbf{E}[\langle \chi, \Theta\rangle] < \infty, \tag{51}$$

and, almost surely, for all $x \in \mathbb{R}_+$,

$$\langle 1_{\{x\}}, \Theta\rangle = 0. \tag{52}$$

Assumption (52) states that a.s. Θ has no atoms, which is needed to show tightness of $\{\bar{\mu}^r(\cdot)\}_{r>0}$ and that fluid limit points have continuous paths a.s. The "no atoms" assumption will be used in the equivalent form

$$\lim_{\kappa\downarrow 0}\mathbf{P}\left(\sup_{x\in\mathbb{R}_+}\langle 1_{[x,\,x+\kappa]},\Theta\rangle < \frac{\epsilon}{4}\right) = 1, \quad \text{for all } \epsilon > 0. \tag{53}$$

The equivalence of (52) and (53) was proved in Gromoll et al. [9, Appendix].

3.3.2. Fluid limit result for strictly supercritical data.

Theorem 3.7. *Let (α,ν) be strictly supercritical data. Consider a sequence of overloaded processor sharing queueing models as defined in §3.3.1, satisfying assumptions (45)–(52) for the data (α,ν). Then, the sequence of fluid-scaled processes $\{\bar{\mu}^r(\cdot)\}_{r>0}$ converges in distribution as $r\to\infty$ to a process $\bar{\mu}^\star(\cdot)$, taking values in $\mathcal{M}_F$ such that $\bar{\mu}^\star(0)$ is equal in distribution to Θ and almost surely $\bar{\mu}^\star(\cdot)$ is a fluid model solution for the data (α,ν).*

We refer to the limiting process $\bar{\mu}^\star(\cdot)$ as the *fluid limit* of the sequence $\{\bar{\mu}^r(\cdot)\}_{r>0}$ of fluid-scaled processes and to Theorem 3.7 as the *fluid limit result*. Section 5 contains the proof of the fluid limit result. This proof is largely an adaptation of the proof of Gromoll et al. [9, Theorem 3.2]. However, if the limiting initial measure is the zero measure, the behavior in the strictly supercritical case diverges from that in the critical case. This is the key issue that is addressed in §5.

The result stated in Jean-Marie and Robert [10, Proposition 4] is a law of large numbers type of result that is consistent with our fluid limit result. The authors of [10] consider a single overloaded processor sharing queue with one job in the system at time zero. To state Jean-Marie and Robert [10, Proposition 4] in terms of our notation, let $\mu(\cdot)$ denote the measure-valued process for the overloaded processor sharing queue defined in [10] and, for each $r>0$, let $\bar{\mu}^r(\cdot)=\mu(r\cdot)/r$. Their proposition states that, under the mild conditions in [10], $\bar{\mu}^r(1)\xrightarrow{w}\bar{\varsigma}(1)$ almost surely as r tends to infinity. To see that this is consistent with our fluid limit result, note that since $\langle 1,\bar{\mu}^r(0)\rangle = 1/r$ for all $r>0$, $\bar{\mu}^r(0)\xrightarrow{w}\mathbf{0}$ almost surely as r tends to infinity. Thus, if conditions (45)–(52) are satisfied, then $\Theta=\mathbf{0}$ and, by Theorem 3.7 here, it follows that

$$\bar{\mu}^r(\cdot)\Longrightarrow\bar{\mu}^\star(\cdot), \quad \text{as } r\to\infty,$$

where $\bar{\mu}^\star(\cdot)=\bar{\varsigma}(\cdot)$ almost surely. In particular, $\bar{\mu}^\star(1)=\bar{\varsigma}(1)$ almost surely.

4. Proofs of properties of supercritical fluid model solutions. This section contains the proofs of the theorems stated but not proved in §3.2. In particular, we prove Theorems 3.2 and 4.1, which imply Theorem 3.1. We also prove Theorem 3.4 and Lemma 4.3 which, together with Gromoll et al. [9, Lemma 4.9], imply Theorem 3.3. Recall that these theorems are concerned with initial measures taking values in the sets $\mathcal{M}_F^c$ and $\mathcal{M}_F^{c,\,p}$. Given $\xi\in\mathcal{M}_F^{c,\,p}$, let

$$h_\xi(x)=\langle 1_{(x,\,\infty)},\xi\rangle, \quad \text{for all } x\in\mathbb{R}_+. \tag{54}$$

Since ξ has no atoms, $h_\xi(\cdot)$ is continuous. Since ξ is not the zero measure, $h_\xi(0)>0$. Moreover, $h_\xi(\cdot)$ is nonnegative and nonincreasing. Let $\mathbf{C}_\searrow^+$ denote the set of all such functions:

$$\mathbf{C}_\searrow^+=\{h\in\mathbf{C}(\mathbb{R}_+)\colon h(0)>0 \text{ and } h(x)\ge h(y)\ge 0 \text{ for all } 0\le x\le y<\infty\}.$$

Denote by $H_\xi(\cdot)$ the antiderivative of $h_\xi(\cdot)$ such that $H_\xi(0)=0$:

$$H_\xi(x)=\int_0^x h_\xi(y)\,dy, \quad \text{for all } x\in\mathbb{R}_+. \tag{55}$$

Then, $H_\xi(\cdot)\in\mathbf{C}^1(\mathbb{R}_+)$. It is easily verified that $H_\xi(x)=\langle x\wedge\chi,\xi\rangle$ for all $x\in\mathbb{R}_+$. Also, recall that, given supercritical data (α,ν), ν_e denotes the excess lifetime probability measure associated with ν (cf. (26)). Let F (resp. F_e) denote the cumulative distribution function associated with the probability measure ν (resp. ν_e). Since ν does not charge the origin, $F(0)=0$. In addition, by (26), $F_e(x)=\beta\int_0^x(1-F(y))dy$, for all $x\ge 0$.

The analysis of the critical fluid model in Gromoll et al. [9] benefited from the behavior of the solutions to certain renewal equations (cf. [9, Equations (4.3) and (4.4)]). For the more general setup of the supercritical fluid model, a factor of ρ must be introduced and the equations then become of a more general Volterra type. Here, we briefly introduce these Volterra equations and summarize the facts required for the supercritical fluid model analysis. Given a locally bounded, Borel-measurable function g: $\mathbb{R}_+ \to \mathbb{R}_+$ and a nondecreasing, right continuous function U: $\mathbb{R}_+ \to \mathbb{R}_+$, let

$$(g * U)(u) = \int_{[0,u]} g(u-s)\,dU(s), \quad \text{for all } u \geq 0.$$

Note that by convention the contribution to the above integral is $g(u)U(0)$ at $s = 0$ whenever $U(0) \neq 0$. Given supercritical data (α, ν) and $h \in \mathbf{C}^+_{\searrow}$, let $H(x) = \int_0^x h(y)\,dy$ for all $x \geq 0$, and consider the convolution equations

$$J(u) = H(u) + \rho(J * F_e)(u), \quad \text{for all } u \geq 0, \tag{56}$$

$$j(u) = h(u) + \rho(j * F_e)(u), \quad \text{for all } u \geq 0. \tag{57}$$

Notice that (after a change of variables from s to $u - s$ in the integrals) the above are indeed linear Volterra equations of the second kind, since $F_e(\cdot)$ is absolutely continuous with density function $f_e(\cdot)$. Since $H(\cdot)$, $h(\cdot)$, and $f_e(\cdot)$ are locally bounded on $\mathbb{R}_+$, each of these equations ((56) and (57)) has a unique locally bounded, Borel-measurable solution. For this, see the proof of Linz [12, Theorem 3.3], which easily generalizes from continuous to locally bounded, Borel-measurable functions. In fact, that proof shows that there is an explicit representation for the unique locally bounded, Borel-measurable solution of each of these equations in terms of the function

$$U_e(u) = \sum_{i=0}^{\infty} \rho^i (F_e^{*i})(u), \quad \text{for all } u \geq 0, \tag{58}$$

where $F_e^{*0}(\cdot) \equiv 1$ and $F_e^{*i}(\cdot) = (F_e^{*(i-1)} * F_e)(\cdot)$, for all $i \in \{1, 2, \ldots\}$. Note that, for each $u \geq 0$, the above series converges since for each $u \geq 0$, $\rho^i(F_e^{*i})(u) \leq (\alpha u)^i / i!$ for all $i \in \{0, 1, \ldots\}$. Then,

$$J(u) = (H * U_e)(u), \quad \text{for all } u \geq 0, \tag{59}$$

is the unique locally bounded, Borel-measurable solution of (56), and

$$j(u) = (h * U_e)(u), \quad \text{for all } u \geq 0, \tag{60}$$

is the unique locally bounded, Borel-measurable solution of (57). Observe that $U_e \in \mathbf{C}(\mathbb{R}_+)$ since $F_e \in \mathbf{C}(\mathbb{R}_+)$. Also, note that $J(0) = 0$, J is strictly increasing, and, since $\rho \geq 1$, $\lim_{u \to \infty} J(u) = \infty$. Moreover, since $U_e \in \mathbf{C}(\mathbb{R}_+)$, $H \in \mathbf{C}^1(\mathbb{R}_+)$, and $H(0) = 0$, it follows that $J \in \mathbf{C}^1(\mathbb{R}_+)$ with $J'(\cdot) = j(\cdot)$.

4.1. Nonzero initial measures: Existence and uniqueness for strictly supercritical data. When the data (α, ν) is strictly supercritical and the initial measure is nonzero, (S.1)–(S.3) imply (C.1)–(C.4) (cf. (21)). Consequently, when the initial measure is nonzero, much of the analysis in Gromoll et al. [9] pertaining to critical data extends to strictly supercritical data as well. In particular, the proof of existence and uniqueness in [9] for nonzero initial measures extends to the strictly supercritical case (cf. Theorem 4.1 below). Similarly, the portion of the proof of [9, Lemma 4.9] which shows that the restriction of Ξ to $\mathcal{M}_F^{c,p}$ is continuous also extends to the strictly supercritical case (cf. Lemma 4.3 below). Here, we outline the key steps for extending these proofs.

Let (α, ν) be strictly supercritical data. Define

$$\mathcal{M}_F^p = \{\xi \in \mathcal{M}_F : \xi \neq \mathbf{0}\}.$$

Our analysis begins with a simple observation: If $\bar{\mu}(\cdot)$ is a fluid model solution for the data (α, ν) such that $\bar{\mu}(0) \in \mathcal{M}_F^p$, then, by (21), $t^* = \infty$. This together with (16) implies that (15) holds. This has some immediate consequences.

Lemma 4.1. *Let (α, ν) be strictly supercritical data. Suppose that $\bar{\mu}(\cdot)$ is a fluid model solution for the data (α, ν) such that $\bar{\mu}(0) \in \mathcal{M}_F^p$. Then, for all $t \geq 0$ and for all $w \in (0, \infty]$,*

$$\langle 1_{(0,w)}, \bar{\mu}(t) \rangle = \langle 1_{(0,w)}(\cdot - \bar{S}(t)), \bar{\mu}(0) \rangle + \alpha \int_0^t \langle 1_{(0,w)}(\cdot - \bar{S}(t) + \bar{S}(s)), \nu \rangle\, ds. \tag{61}$$

The equations in (61) (one for each $t \geq 0$ and $w \in (0, \infty]$) have an intuitive interpretation in terms of the fluid model dynamics. To see this, note that for $0 \leq s \leq t$, $\bar{S}(t) - \bar{S}(s)$ represents the net shift to the left due to service in the time interval $(s, t]$. Thus, the integrand in the second term on the right-hand side of (4.8) accounts for the contribution to the left side at time t due to the fluid that arrived at time $0 \le s \le t$. Similarly, the first term on the right-hand side of (4.8) accounts for the contribution due to fluid that was in the system at time 0.

Note that (61) also holds for critical data as well, at least for $0 \le t < t^*$ (cf. Gromoll et al. [9, Lemma 4.3]). In fact, the proof of Lemma 4.1 is almost identical to the proof of [9, Lemma 4.3] and thus it is omitted. The interested reader can readily verify that the necessary ingredients are continuity of $\bar{\mu}(\cdot)$ and (15) together with $t^* = \infty$, which is guaranteed by (21). Specifically, it is straightforward to check that the property that $\bar{\mu}(0)$ has no atoms, which is a condition of [9, Lemma 4.3], is not used in its proof. In fact, as a consequence of Lemma 4.1, we have the following result that the initial measure has no atoms.

Proposition 4.1. *Let (α, ν) be strictly supercritical data. Suppose that $\bar{\mu}(\cdot)$ is a fluid model solution for the data (α, ν) such that $\bar{\mu}(0) \in \mathcal{M}_F^p$. Then, $\bar{\mu}(0) \in \mathcal{M}_F^{c,p}$.*

Proof. Fix strictly supercritical data (α, ν). Suppose that $\bar{\mu}(\cdot)$ is a fluid model solution for the data (α, ν) such that $\bar{\mu}(0) \in \mathcal{M}_F^p$. By (S.2), $\bar{\mu}(0)$ does not have an atom at the origin. So it suffices to show that $\bar{\mu}(0)$ has no atoms on the positive real line. For this, we will show that, if $\bar{\mu}(0)$ has an atom at $a \in (0, \infty)$, then $\bar{Z}(\cdot)$ has a discontinuity at time t_a, the time at which this atom departs the system. This will violate (S.1) and provide a contradiction, thereby implying that no such a exists.

For this, we first need to show that an atom that starts at $a \in (0, \infty)$ does eventually depart the system, i.e., that there exists a finite time t_a such that $\bar{S}(t_a) = a$. This follows from the fact that

$$\lim_{t \to \infty} \bar{S}(t) = \infty. \tag{62}$$

To verify (62), note that by (61) with $w = \infty$ and (S.2), for each $t \ge 0$,

$$\bar{Z}(t) = \langle 1_{(0,\infty)}(\cdot - \bar{S}(t)), \bar{\mu}(0) \rangle + \alpha \int_0^t \langle 1_{(0,\infty)}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \rangle \, ds. \tag{63}$$

Equation (63), together with (21), implies that

$$0 < \bar{Z}(t) \le \bar{Z}(0) + \alpha t, \quad \text{for all } t \ge 0.$$

Therefore, for all $t \ge 0$,

$$\bar{S}(t) = \int_0^t \frac{1}{\bar{Z}(s)} \, ds \ge \frac{1}{\alpha} \log\left(\frac{\bar{Z}(0) + \alpha t}{\bar{Z}(0)}\right),$$

and (62) follows.

Suppose that $a \in (0, \infty)$ is such that $\bar{\mu}(0)$ has an atom at a, and let $t_a > 0$ be such that $\bar{S}(t_a) = a$. Clearly, the first term on the right side of (63) has a downward jump at time t_a. Thus, to complete the proof that $\bar{Z}(\cdot)$ has a discontinuity at $t = t_a$, it suffices to show that the second term on the right side of (63) is continuous for $t \in [0, \infty)$. For this, note that for each fixed $t \ge 0$ and $s \in [0, t]$ such that ν does not have an atom at $\bar{S}(t) - \bar{S}(s)$,

$$\lim_{u \to t} \langle 1_{(0,\infty)}(\cdot - (\bar{S}(u) - \bar{S}(s))), \nu \rangle = \langle 1_{(0,\infty)}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \rangle.$$

Since $\bar{S}(\cdot)$ is strictly increasing and ν has at most countably many atoms, the above holds except perhaps for countably many values of $s \in [0, t]$, which comprise a set of Lebesgue measure zero (possibly depending on t). So, using bounded convergence and the fact that the integrand is bounded, continuity of the second term on the right side of (63) for $t \in [0, \infty)$ follows. ☐

Remark 4.1. Proving a version of Proposition 4.1 for critical data is possible, but slightly more delicate. This is so because the proof of Proposition 4.1 requires (21), i.e., $t^* = \infty$. By following the arguments in Gromoll et al. [9], it can be verified that $t^* = \infty$ under the condition $\bar{\mu}(0) \in \mathcal{M}_F^p$. In particular, it is easy to verify that the condition $\bar{\mu}(0) \in \mathcal{M}_F^p$ is all that is required in the proof of Lemma 4.4 in [9]. Specifically, the proof of [9, Lemma 4.4] relies on Lemma 4.3 there and convolution equation arguments, neither of which requires $\bar{\mu}(0)$ to have no atoms. Indeed, one can justify the conclusion $\bar{T}(\cdot) = (H_\xi * U_e)(\cdot)$ using the facts that $\bar{T}(\cdot)$ is continuously differentiable with $\bar{T}'(\cdot) = (h_\xi * U_e)(\cdot)$ and $H_\xi(\cdot) = \int_0^\cdot h_\xi(y)\, dy$, together with Fubini's theorem to justify interchanging the order of integration.

Corollary 4.1. *Let (α, ν) be strictly supercritical data. Suppose that $\bar{\mu}(\cdot)$ is a fluid model solution for the data (α, ν) such that $\bar{\mu}(0) \in \mathcal{M}_F^p$. Then, $\bar{\mu}(t) \in \mathcal{M}_F^{c,p}$ for all $t \ge 0$.*

Proof. Fix strictly supercritical data (α, ν). Suppose that $\bar{\mu}(\cdot)$ is a fluid model solution for the data (α, ν) such that $\bar{\mu}(0) \in \mathcal{M}_F^p$. By (21), $\bar{\mu}(s) \in \mathcal{M}_F^p$ for all $s \geq 0$. By Lemma 3.1, for each $s \geq 0$, $(\tau_s \bar{\mu})(\cdot)$ is a fluid model solution for the data (α, ν). Note that, for each $s \geq 0$, $(\tau_s \bar{\mu})(0) = \bar{\mu}(s)$. Thus, for each $s \geq 0$, $(\tau_s \bar{\mu})(0) \in \mathcal{M}_F^p$. Consequently, by Proposition 4.1, $\bar{\mu}(s) = (\tau_s \bar{\mu})(0) \in \mathcal{M}_F^{c,p}$ for each $s \geq 0$. □

Remark 4.2. In light of Remark 4.1, a similar proof to that given above yields that the conclusion of Corollary 4.1 also holds for critical data.

Given strictly supercritical data (α, ν) and $\xi \in \mathcal{M}_F^{c,p}$, suppose that $\bar{\mu}(\cdot)$ is a fluid model solution for the data (α, ν) such that $\bar{\mu}(0) = \xi$. It turns out that (63) can also be used to relate $\bar{Z}(\cdot)$ to a time changed solution of a certain convolution equation. To see this, observe that $\bar{S}(\cdot)$ is continuous and strictly increasing (since $\bar{Z}(\cdot)$ is continuous and strictly positive). Moreover, by (62), $\bar{S}(\cdot)$ maps onto $[0, \infty)$. Thus, the continuous inverse of $\bar{S}(\cdot)$ is given by

$$\bar{T}(u) = \bar{S}^{-1}(u) = \inf\{t \geq 0: \bar{S}(t) > u\}, \quad \text{for all } u \geq 0. \tag{64}$$

In (64), the superscript -1 denotes the functional inverse. The function $\bar{T}(\cdot)$ has a natural interpretation in terms of the fluid model dynamics. For $u \geq 0$, $\bar{T}(u)$ represents the amount of time that fluid that is at u at time 0 spends in the system, i.e., the fluid analogue of sojourn time. Note that, since $\bar{\mu}(\cdot)$ is continuous and never zero, $\bar{S}(\cdot) \in \mathbf{C}^1([0, \infty))$ with $\bar{S}'(\cdot) = 1/\bar{Z}(\cdot)$. This means that $\bar{T}(\cdot) \in \mathbf{C}^1([0, \infty))$ with

$$\bar{T}'(u) = \frac{1}{\bar{S}'(\bar{T}(u))}, \quad \text{for all } u \geq 0. \tag{65}$$

In other words, $\bar{T}'(\bar{S}(t)) = \bar{Z}(t)$ for all $t \geq 0$. Using this fact, together with (63), yields an equation for $\bar{T}'(\bar{S}(\cdot))$. One can use this equation to show that $\bar{T}'(\cdot) = j_\xi(\cdot)$, where $j_\xi(\cdot)$ is the unique locally bounded solution of the convolution equation (57) for $h(\cdot) = h_\xi(\cdot)$ with h_ξ defined by (54). Indeed, we have the following lemma.

Lemma 4.2. *Let (α, ν) be strictly supercritical data and let $\xi \in \mathcal{M}_F^{c,p}$. Suppose that $\bar{\mu}(\cdot)$ is a fluid model solution for the data (α, ν) such that $\bar{\mu}(0) = \xi$. If $\bar{T}(\cdot)$ is defined by (64), then*

$$\bar{T}(u) = (H_\xi * U_e)(u) \quad \textit{and} \quad \bar{T}'(u) = (h_\xi * U_e)(u), \qquad \textit{for all } u \geq 0.$$

The analog of Lemma 4.2 for critical data was proved in Gromoll et al. [9] (cf. [9, Lemma 4.4]). In fact, the proof of Lemma 4.2 is nearly identical to the proof of [9, Lemma 4.4], and therefore it is not included here. The main difference is that for strictly supercritical data, a factor of ρ is picked up $\alpha(1 - F(x)) = \rho f_e(x)$, for all $x \geq 0$.

Theorem 4.1. *Let (α, ν) be strictly supercritical data and let $\xi \in \mathcal{M}_F^{c,p}$. A fluid model solution $\bar{\mu}(\cdot)$ for the data (α, ν) such that $\bar{\mu}(0) = \xi$ exists and is unique.*

Proof. Suppose that $\bar{\mu}(\cdot)$ is a fluid model solution for the strictly supercritical data (α, ν) such that $\bar{\mu}(0) = \xi$. By Lemma 4.1 and (S.2), $\langle 1_{[0,w)}, \bar{\mu}(t) \rangle$ is uniquely determined by $\bar{S}(\cdot)$, ξ, and (α, ν) for each $w \in (0, \infty]$ and for each $t \in [0, \infty)$. Since intervals of the form $[0, w)$, $w \in (0, \infty]$, generate the Borel σ-algebra on $\mathbb{R}_+$, this uniquely determines $\bar{\mu}(t)$ for each $t \in [0, \infty)$. By (64) and Lemma 4.2, $\bar{S}(\cdot)$ is the inverse of $(H_\xi * U_e)(\cdot)$. Since $(H_\xi * U_e)(\cdot)$, is determined by ξ and (α, ν), the function $\bar{S}(\cdot)$, and therefore the fluid model solution $\bar{\mu}(\cdot)$, is uniquely determined by ξ and (α, ν).

To prove existence, one can mimic the construction of a fluid model solution in §4.2 of Gromoll et al. [9], except that $U_e(\cdot)$ is defined by (58). Such a $\bar{\mu}(\cdot)$ satisfies (S.1), (S.2), and $\bar{\mu}(0) = \xi$. Moreover, $\bar{\mu}(t) \neq \mathbf{0}$ for all $t \geq 0$ and (15) holds for this $\bar{\mu}(\cdot)$. Therefore, (S.3) holds and existence follows. □

Let (α, ν) be supercritical data and let Ξ_p be the restriction of Ξ to $\mathcal{M}_F^{c,p}$. Under the condition that $\rho = 1$, it was proved in Gromoll et al. [9] that this mapping is continuous (cf. the proof of [9, Lemma 4.9]). It is straightforward to check that this proof generalizes to the situation in which $\rho > 1$. Thus, we obtain the following lemma.

Lemma 4.3. *Let (α, ν) be strictly supercritical data. Then, Ξ_p is continuous.*

4.2. The zero initial measure: Existence for strictly supercritical data.

Lemma 4.4. *Let (α, ν) be strictly supercritical data. Then, $\bar{\varsigma}(\cdot)$ defined by (28), (29), and (31) is a fluid model solution for the data (α, ν).*

Proof. Clearly, $\bar{\varsigma}(\cdot)$ satisfies (S.1) and (S.2). Therefore, it suffices to show that $\bar{\varsigma}(\cdot)$ satisfies (S.3), i.e., that $\bar{\varsigma}(\cdot)$ is a solution of (16). Since $\bar{\varsigma}(t) \neq \mathbf{0}$ for all $t > 0$, $\bar{\varsigma}(\cdot)$ satisfies (16) if and only if, for all $g \in \mathscr{C}$ and $t > 0$,

$$t\langle g, \varsigma\rangle = -\frac{t}{m}\langle g', \varsigma\rangle + \alpha t\langle g, \nu\rangle, \tag{66}$$

which is equivalent to

$$\frac{1}{m}\langle g', \varsigma\rangle + \langle g, \varsigma\rangle = \alpha\langle g, \nu\rangle, \quad \text{for all } g \in \mathscr{C}. \tag{67}$$

In order to verify (67), fix $g \in \mathscr{C}$, combine the two integrals on the left side of (67) into one, use the definition of ς and of $p(\cdot)$, interchange the order of integration, note that for each $y \in \mathbb{R}_+$, $(g'(\cdot) + mg(\cdot))\exp(m(\cdot - y))$ is a perfect derivative, and use the fact that $g(0) = 0$. □

Lemma 4.4 establishes the existence of a fluid model solution for the zero initial measure. Thus, in order to prove Theorem 3.2, it suffices to prove uniqueness. For this, we will need to use an order preservation property and a weak continuity property for fluid model solutions. These are stated and proved in §4.3 and §4.10, respectively. Those statements and proofs require versions of (61) and (63) that hold for $\bar{\varsigma}(\cdot)$; these are provided by Lemma 4.5, and Corollary 4.2, respectively.

Lemma 4.5. *Let (α, ν) be strictly supercritical data. Then, for all $t > 0$ and $x \in \mathbb{R}_+$,*

$$\langle 1_{(x,\infty)}, \bar{\varsigma}(t)\rangle = \rho\int_{(0,t]} f_e\left(x + \frac{1}{m}\log\left(\frac{t}{s}\right)\right) ds. \tag{68}$$

Proof. We begin with a simple calculation. By using the definiton of $p(\cdot)$, interchanging the order of integration, and performing a change of variables, we obtain, for all $x \in \mathbb{R}_+$,

$$\int_{(x,\infty)} p(y)\,dy = \alpha\int_{(x,\infty)}\int_{(0,z-x)} \exp(-my)\,dy\,\nu(dz).$$

After another interchange of the order of integration, a use of the definition of $f_e(\cdot)$, and a change of variables, for all $x \in \mathbb{R}_+$ we obtain

$$\int_{(x,\infty)} p(y)\,dy = \rho\int_{(x,\infty)} f_e(y)\exp(-m(y-x))\,dy. \tag{69}$$

So, from the definiton of $\bar{\varsigma}(\cdot)$, we see that, for each $t > 0$ and $x \in \mathbb{R}_+$,

$$\langle 1_{(x,\infty)}, \bar{\varsigma}(t)\rangle = mt\rho\int_{(x,\infty)} f_e(y)\exp(-m(y-x))\,dy.$$

Using the change of variables $s = t\exp(-m(y-x))$ in the above integral proves (68). □

By taking $x = 0$ in (68), using the definition of $\bar{\varsigma}(\cdot)$, and using (30), we obtain the following corollary.

Corollary 4.2. *Let (α, ν) be strictly supercritical data. Then,*

$$mt = \rho\int_{(0,t]} f_e\left(\frac{1}{m}\log\left(\frac{t}{s}\right)\right) ds, \quad \textit{for all } t > 0. \tag{70}$$

4.3. An order preservation property. In this section, we prove an order preservation property (cf. Theorem 4.2) that turns out to be a key tool for analyzing the supercritical fluid model. In particular, it is used in §4.4 to prove a conservation law (cf. Lemma 4.9), which is used in turn to prove a weak continuity at zero property (cf. Lemma 4.10). The weak continuity at zero property is used to prove uniqueness for Theorem 3.2. Moreover, some of the tools developed for proving the order preservation property (cf. Lemma 4.8) are also used in §4.5 to prove a continuity property (cf. Theorem 4.3), which is the key ingredient in the proof of Theorem 3.4.

To formulate the order preservation property, we fix supercritical data (α, ν) and investigate the effect of the fluid model dynamics on the following partial ordering on $\mathscr{M}_F$.

Definition 4.1. Suppose that $\zeta_1, \zeta_2 \in \mathscr{M}_F$. Then, ζ_2 dominates ζ_1 if

$$\langle 1_{(x,\infty)}, \zeta_1\rangle \le \langle 1_{(x,\infty)}, \zeta_2\rangle, \quad \text{for all } x \in \mathbb{R}_+,$$

in which case we write $\zeta_1 \preceq \zeta_2$.

Since the processor sharing queue is a discrete event system, it is straightforward to verify that the processor sharing queueing dynamics preserve this partial ordering. Specifically, if $\mu_1(\cdot)$ and $\mu_2(\cdot)$ are measure-valued state descriptors for a processor sharing queue such that $\mu_1(0) \preceq \mu_2(0)$, then it is not difficult to verify that $\mu_1(t) \preceq \mu_2(t)$ for all $t \geq 0$. The aim of this section is to prove an analogous result for the fluid model dynamics. In particular, the main objective of this section is to prove Theorem 4.2 below, which implies that, under appropriate conditions, the fluid model dynamics preserve this partial ordering. In addition, Theorem 4.2 implies that if a fluid model solution $\bar{\mu}(\cdot)$ has a nonzero initial measure, then $\bar{\mu}(t)$ dominates $\bar{\varsigma}(t)$ for all $t \geq 0$. For this, we have adopted the convention that

$$\bar{\varsigma}(\cdot) \equiv \mathbf{0}, \quad \text{if } \rho = 1.$$

In order to simplify the statement of Theorem 4.2, we introduce some notation. For $\xi \in \mathcal{M}_F^{c,p}$, let $\bar{\mu}_\xi(\cdot)$ denote the unique fluid model solution for the supercritical data (α, ν) such that $\bar{\mu}_\xi(0) = \xi$. Given $\xi \in \mathcal{M}_F^{c,p}$, for all $t \geq 0$, let

$$\bar{Z}_\xi(t) = \langle 1, \bar{\mu}_\xi(t) \rangle, \qquad \bar{S}_\xi(t) = \int_0^t \frac{1}{\bar{Z}_\xi(s)}\, ds, \qquad \bar{T}_\xi(t) = \bar{S}_\xi^{-1}(t), \qquad \bar{W}_\xi(t) = \langle \chi, \bar{\mu}_\xi(t) \rangle,$$

and

$$\bar{M}_\xi(t, x) = \langle 1_{(x,\infty)}, \bar{\mu}_\xi(t) \rangle, \quad \text{for all } x \in \mathbb{R}_+.$$

In the above, we have used $\xi \in \mathcal{M}_F^{c,p}$, (21) here, and Gromoll et al. [9, Theorem 3.1] to eliminate $\varphi(\cdot)$ in the definition of $\bar{S}_\xi(\cdot)$. In this section, we will frequently consider $\xi_1, \xi_2 \in \mathcal{M}_F^{c,p}$, and in future sections, we will frequently consider $\{\xi_n, n \in \mathbb{N}\} \subset \mathcal{M}_F^{c,p}$. To avoid cluttering the notation, for each $n \in \mathbb{N}$, we write

$$\bar{\mu}_n(\cdot) = \bar{\mu}_{\xi_n}(\cdot).$$

Similarly, for each $n \in \mathbb{N}$ and for all $t \geq 0$, we let $\bar{Z}_n(\cdot) = \bar{Z}_{\xi_n}(\cdot)$, $\bar{S}_n(\cdot) = \bar{S}_{\xi_n}(\cdot)$, $\bar{T}_n(\cdot) = \bar{T}_{\xi_n}(\cdot)$, $\bar{W}_n(\cdot) = \bar{W}_{\xi_n}(\cdot)$, and $\bar{M}_n(\cdot, \cdot) = \bar{M}_{\xi_n}(\cdot, \cdot)$. As noted above, we have adopted the convention that if $\rho = 1$, then $\bar{\varsigma}(\cdot) \equiv \mathbf{0}$. In addition, we set

$$m = 0, \quad \text{if } \rho = 1. \tag{71}$$

Theorem 4.2. *Let $\xi_1, \xi_2 \in \mathcal{M}_F^{c,p}$ be such that $\xi_1 \preceq \xi_2$. Then, for all $t \geq 0$,*

$$\bar{\varsigma}(t) \preceq \bar{\mu}_1(t) \preceq \bar{\mu}_2(t).$$

In order to prove Theorem 4.2, we must verify that, for each $t \geq 0$,

$$\langle 1_{(x,\infty)}, \bar{\varsigma}(t) \rangle \leq \bar{M}_1(t, x) \leq \bar{M}_2(t, x), \quad \text{for all } x \in \mathbb{R}_+. \tag{72}$$

Note that if $\rho > 1$, then for each $t \geq 0$ and $x \in \mathbb{R}_+$, (68) expresses $\langle 1_{(x,\infty)}, \bar{\varsigma}(t) \rangle$ in terms of the data (α, ν). (Recall that m and ρ are determined by (α, ν).) The next lemma gives an analogue of (68) for nonzero initial measures that is valid in the case of both critical data and strictly supercritical data.

Lemma 4.6. *Let $\xi \in \mathcal{M}_F^{c,p}$. Then, $\bar{M}_\xi(t, x)$ is jointly continuous as a function of $(t, x) \in [0, \infty) \times \mathbb{R}_+$ and*

$$\bar{M}_\xi(t, x) = \bar{M}_\xi(0, x + \bar{S}_\xi(t)) + \rho \int_0^t f_e(x + \bar{S}_\xi(t) - \bar{S}_\xi(s))\, ds, \quad \textit{for all } t \geq 0 \textit{ and } x \in \mathbb{R}_+. \tag{73}$$

Proof. Fix $\xi \in \mathcal{M}_F^{c,p}$. To verify (73), fix $t \geq 0$. Note that by definition $\bar{M}_\xi(t, x) = \bar{Z}_\xi(t) - \langle 1_{[0,x]}, \bar{\mu}_\xi(t) \rangle$, for all $x \in \mathbb{R}_+$. By reasoning similar to that in the proof of Gromoll et al. [9, Proposition 4.6] and since $t^* = \infty$ for critical data, (61) holds with $1_{[0,w]}$ in place of $1_{(0,w)}$ for all supercritical data. Setting $w = \infty$ and $w = x$ and subtracting yields (73).

To verify the joint continuity of $\bar{M}_\xi$, we will verify the joint continuity of each term on the right side of (73). Note that since ξ has no atoms, $\bar{M}_\xi(0, x)$ is a continuous function of $x \in \mathbb{R}_+$. Also, since $\bar{\mu}_\xi(\cdot)$ is continuous and $\bar{Z}_\xi(\cdot)$ is never zero (cf. Proposition 4.1), $\bar{S}_\xi(\cdot)$ is continuous. Therefore, $\bar{M}_\xi(0, x + \bar{S}(t))$ is jointly continuous as a function of $(t, x) \in [0, \infty) \times \mathbb{R}_+$. To verify the joint continuity of the second term on the right side of (73), note that for each $t \geq 0$, $s \in [0, t]$, and $x \in \mathbb{R}_+$ such that ν does not have an atom at $x + \bar{S}_\xi(t) - \bar{S}_\xi(s)$,

$$\lim_{(u,y)\to(t,x)} f_e(y + \bar{S}_\xi(u) - \bar{S}_\xi(s)) = f_e(x + \bar{S}_\xi(t) - \bar{S}_\xi(s)).$$

Since $\bar{S}(\cdot)$ is strictly increasing and ν has at most countably many atoms, the above holds except perhaps for countably many values of $s \in [0, t]$, which comprise a set of Lebesgue measure zero (possibly depending on x and t). So, using bounded convergence and the fact that the integrand is bounded, it follows that the second term on the right side of (73) is jointly continuous as a function of $(t, x) \in [0, \infty) \times \mathbb{R}_+$. □

We will use (68) and (73) to prove Lemma 4.7 below. This lemma implies that, if in addition to the conditions of Theorem 4.2, we also know that

$$mt \leq \bar{Z}_1(t) \leq \bar{Z}_2(t), \quad \text{for all } t \geq 0, \tag{74}$$

then $\bar{\varsigma}(t) \preceq \bar{\mu}_1(t) \preceq \bar{\mu}_2(t)$ for all $t \geq 0$. The conditions of Lemma 4.7 are more easily verified than (72). To see this, note that by (C.2) and (S.2), (74) is the $x = 0$ instance of (72).

Lemma 4.7. *Let $\xi_1, \xi_2 \in \mathcal{M}_F^{c,p}$. Then, $\bar{\varsigma}(t) \preceq \bar{\mu}_1(t) \preceq \bar{\mu}_2(t)$ for all $t \geq 0$ if and only if $\xi_1 \preceq \xi_2$ and $mt \leq \bar{Z}_1(t) \leq \bar{Z}_2(t)$, for all $t \geq 0$.*

Proof. Fix ξ_1, $\xi_2 \in \mathcal{M}_F^{c,p}$. The "only if" direction follows by taking $t = 0$ for all x, and then $x = 0$ for all t in (72) and using (C.2) and (S.2). Therefore, it suffices to prove the "if" direction. For this, we need to verify (72) assuming that $\xi_1 \preceq \xi_2$ and $mt \leq \bar{Z}_1(t) \leq \bar{Z}_2(t)$ for all $t \geq 0$. By (73), we have, for $i = 1, 2$,

$$\bar{M}_i(t, x) = \bar{M}_i(0, x + \bar{S}_i(t)) + \rho \int_0^t f_e(x + \bar{S}_i(t) - \bar{S}_i(s))\, ds, \quad \text{for all } t \geq 0 \text{ and } x \in \mathbb{R}_+. \tag{75}$$

Since $mu \leq \bar{Z}_1(u) \leq \bar{Z}_2(u)$ for all $u \geq 0$,

$$\bar{S}_1(t) - \bar{S}_1(s) \geq \bar{S}_2(t) - \bar{S}_2(s), \quad \text{for all } 0 \leq s \leq t < \infty. \tag{76}$$

This, together with (75), the fact that $\xi_1 \preceq \xi_2$, the fact that $\bar{M}_i(0, \cdot)$ is nonincreasing for $i = 1, 2$, and the fact that $f_e(\cdot)$ is nonincreasing implies that $\bar{\mu}_1(t) \preceq \bar{\mu}_2(t)$ for all $t \geq 0$. If the data is critical, then $\bar{\varsigma}(\cdot) \equiv \mathbf{0}$, and the proof is complete. Otherwise, the data is strictly supercritical, and then, since $mu \leq \bar{Z}_1(u)$ for all $u \geq 0$, we have

$$\frac{1}{m} \log\left(\frac{t}{s}\right) \geq \bar{S}_1(t) - \bar{S}_1(s), \quad \text{for all } 0 < s \leq t < \infty. \tag{77}$$

This, together with (68) and (75), the fact that $\bar{M}_1(0, \cdot)$ is nonnegative, and the fact that $f_e(\cdot)$ is nonincreasing implies that, if the data is strictly supercritical, then for all $t > 0$, $\bar{\varsigma}(t) \preceq \bar{\mu}_1(t)$. Since $\bar{\varsigma}(0) = \mathbf{0}$, this also holds for $t = 0$, and the proof is complete. □

By Lemma 4.7, in order to prove Theorem 4.2, it suffices to verify that, under the conditions of Theorem 4.2, (74) holds. Let $\xi \in \mathcal{M}_F^{c,p}$. Then, by (65) if the data is strictly supercritical, or by Gromoll et al. [9, Equation (4.20) and Lemma 4.4] if the data is critical,

$$\bar{T}'_\xi(\bar{S}_\xi(\cdot)) = \bar{Z}_\xi(\cdot). \tag{78}$$

Also by Lemma 4.2 here and Gromoll et al. [9, Lemma 4.4], $\bar{T}'_\xi(\cdot)$ is the unique locally bounded, Borel-measurable solution of (57) for $h(\cdot) = h_\xi(\cdot)$. In addition, notice that, for $\xi_1, \xi_2 \in \mathcal{M}_F^{c,p}$, the condition $\xi_1 \preceq \xi_2$ is equivalent to $h_{\xi_1}(\cdot) \leq h_{\xi_2}(\cdot)$. This suggests that one needs to determine how suitable time changes of the solutions to the convolution equations (57) are ordered with respect to each other for suitably ordered choices of the functions h.

We begin by defining the time change $K(\cdot)$ for a given $h(\cdot)$. Let $h \in \mathbf{C}^+_\searrow$ and let $j(\cdot)$ be given by (60). In particular, $j(\cdot)$ is continuous and strictly positive. Also, for each $x \in \mathbb{R}_+$, define $H(x) = \int_0^x h(y)\, dy$ and let $J(\cdot)$ be given by (59). Then, as noted immediately below (60), $J \in \mathbf{C}^1(\mathbb{R}_+)$, $J(\cdot)$ is strictly increasing, and $\lim_{u \to \infty} J(u) = \infty$. We wish to define a time change $K(\cdot)$ such that $j(K(\cdot))$ plays the role of $\bar{Z}(\cdot)$. For this, we recall that $\bar{S}(\cdot)$ is the functional inverse of $\bar{T}(\cdot)$. Thus, we define

$$K(t) = \inf\{u \geq 0\colon J(u) > t\}, \quad \text{for all } t \geq 0, \tag{79}$$

which is the functional inverse of $J(\cdot)$. Let

$$L(t) = j(K(t)), \quad \text{for all } t \geq 0. \tag{80}$$

The next proposition implies that for $\xi \in \mathcal{M}_F^{c,p}$ and $h = h_\xi$, $L(\cdot)$ is the desired time change of $j(\cdot)$.

Proposition 4.2. *Let $\xi \in \mathcal{M}_F^{c,p}$. Then, $\bar{Z}_\xi(t) = L(t)$ for all $t \geq 0$, where $L(\cdot)$ is defined by (80) for $h(\cdot) = h_\xi(\cdot)$.*

Proof. Fix $\xi \in \mathcal{M}_F^{c,p}$. Let $j(\cdot)$ be given by (60) with $h(\cdot) = h_\xi(\cdot)$. By Lemma 4.2 here and Gromoll et al. [9, Lemma 4.4], $j(\cdot) = \overline{T}'_\xi(\cdot)$ and $J(\cdot) = \overline{T}_\xi(\cdot)$. Hence $K(\cdot) = \overline{S}_\xi(\cdot)$. In particular, $L(\cdot) = \overline{T}'_\xi(\overline{S}_\xi(\cdot))$. This together with (78) completes the proof. □

We wish to show that if $h_1, h_2 \in \mathbf{C}^+_\searrow$ and $h_1(x) \le h_2(x)$ for all $x \in \mathbb{R}_+$, then $mt \le L_1(t) \le L_2(t)$ for all $t \ge 0$, where, for $i = 1, 2$, $L_i(\cdot)$ is defined by (80) with $h = h_i$. We will in fact verify this conclusion under the more restrictive condition $0 < h_1(x) < h_2(x)$ for all $x \in \mathbb{R}_+$ (cf. Lemma 4.8 below). We will then use this to prove the desired result. For this, we need to establish some facts about $L(\cdot)$. Since K is the functional inverse of J, $J' = j$, and j is strictly positive, it follows that $K \in \mathbf{C}^1(\mathbb{R}_+)$ and

$$K'(t) = \frac{1}{L(t)}, \quad \text{for all } t \ge 0. \tag{81}$$

Also, by (57), symmetry of convolution densities, the change of variables $v = K(s)$, and (80)–(81), we have

$$\begin{aligned} L(t) = j(K(t)) &= h(K(t)) + \rho(j * F_e)(K(t)) \\ &= h(K(t)) + \rho \int_0^{K(t)} j(K(t)-v) f_e(v)\, dv \\ &= h(K(t)) + \rho \int_0^{K(t)} f_e(K(t)-v) j(v)\, dv \\ &= h(K(t)) + \rho \int_0^{t} f_e(K(t)-K(s))\, ds. \end{aligned} \tag{82}$$

Lemma 4.8.
(i) *Let $h \in \mathbf{C}^+_\searrow$ be such that $0 < h(x)$ for all $x \in \mathbb{R}_+$. Then, for all $t \ge 0$,*

$$mt \le L(t),$$

where $L(\cdot)$ is given by (80).
(ii) *Let $h_1, h_2 \in \mathbf{C}^+_\searrow$ be such that $h_1(x) < h_2(x)$ for all $x \in \mathbb{R}_+$. Then, for all $t \ge 0$,*

$$L_1(t) \le L_2(t),$$

where, for $i = 1, 2$, $L_i(\cdot)$ is given by (80) with $h(\cdot) = h_i(\cdot)$.

Proof. For critical data, (i) is trivial since $m = 0$ in that case. To prove (i) for strictly supercritical data, let $T = \inf\{t \ge 0\colon mt > L(t)\}$. Suppose that $T < \infty$. Then, it follows that $T > 0$,

$$mu \le L(u) \quad \text{for all } u \in [0, T], \tag{83}$$

and

$$mT = L(T). \tag{84}$$

To verify this, use the fact that $L(\cdot)$ is continuous (since it is a composition of continuous functions) and the fact that $0 < h(0) = L(0)$. By (81) and (83),

$$\frac{1}{m}\log\left(\frac{T}{s}\right) = \int_s^T \frac{1}{mu}\, du \ge \int_s^T \frac{1}{L(u)}\, du = K(T) - K(s), \quad s \in (0, T].$$

Using this inequality and the facts that $f_e(\cdot)$ is nonincreasing and $0 < h(\cdot)$ gives

$$0 < h(K(T)) \qquad \text{and} \qquad f_e\left(\frac{1}{m}\log\left(\frac{T}{s}\right)\right) \le f_e(K(T) - K(s)), \quad s \in (0, T].$$

This, together with (70) and (82), implies that $mT < L(T)$. But this contradicts (84), and therefore $T = \infty$. In particular, (i) holds for strictly supercritical data.

The proof of (ii) is similar. In particular, let $T = \inf\{t \ge 0\colon L_1(t) > L_2(t)\}$ and suppose that $T < \infty$, which implies $L(T_1) = L(T_2)$ and $L_1(u) \le L_2(u)$ for $u \in [0, T]$. To obtain the contradication $L_1(T) < L_2(T)$, proceed in a manner similar to the above, using (82), $L_1(u) \le L_2(u)$ for $u \in [0, T]$, and $h_1(x) < h_2(x)$ for all $x \in \mathbb{R}_+$. □

Proof of Theorem 4.2. By Lemma 4.7, it suffices to show that (74) holds. Note that, since $\xi_1 \preceq \xi_2$, it follows that $0 \le h_{\xi_1}(\cdot) \le h_{\xi_2}(\cdot)$. In order to use Lemma 4.8, we fix sequences $\{\xi_{i,n}, n \in \mathbb{N}\} \subset \mathcal{M}_F^{c,p}$, $i = 1, 2$, such that $\xi_{i,n} \xrightarrow{w} \xi_i$ as $n \to \infty$ and

$$0 < h_{\xi_{1,n}}(\cdot) < h_{\xi_{2,n}}(\cdot), \quad \text{for each } n \in \mathbb{N}.$$

By Lemma 4.8 and Proposition 4.2, for each $n \in \mathbb{N}$, $mt \le \bar{Z}_{1,n}(t) \le \bar{Z}_{2,n}(t)$ for all $t \ge 0$, where, for $i = 1, 2$ and for each $n \in \mathbb{N}$, $\bar{\mu}_{i,n}(\cdot)$ is the unique fluid model solution with $\bar{\mu}_{i,n}(0) = \xi_{i,n}$ and $\bar{Z}_{i,n}(\cdot) = \langle 1, \bar{\mu}_{i,n}(\cdot)\rangle$. By letting n tend to infinity and using the continuity of Ξ_p (cf. Lemma 4.3 here or Gromoll et al. [9, Lemma 4.9] for the critical case), it follows that $mt \le \bar{Z}_1(t) \le \bar{Z}_2(t)$ for all $t \ge 0$. □

4.4. Consequences of the order preservation property. Presently, we are in the process of developing the machinery to prove the uniqueness in Theorem 3.2, and to prove Theorem 3.4. In §4.3, the order preservation property was stated and proved (cf. Theorem 4.2). Here we use the order preservation property to prove a conservation law (cf. Lemma 4.9) and weak continuity at zero property (cf. Lemma 4.10). The weak continuity at zero property is used in §4.4.3 to prove the uniqueness in Theorem 3.2.

Recall that, in §3.2.6, Theorem 3.5 was proved as a simple consequence of Theorem 3.4. In turn, Theorem 3.6 was proved in §3.2.7 as a consequence of Theorem 3.5, and simple properties of the measure ς. So, Theorem 3.4 is the cornerstone of the three results. The proof of Theorem 3.4 given in §4.5 requires substantial effort. However, if one is willing to impose conditions that control the behavior of the fluid analogue of the workload process, then weak versions of Theorems 3.5 and 3.6 can be obtained as direct consequences of the weak continuity at zero property (cf. Lemmas 4.11 and 4.12). These weak versions are proved in §4.4.4.

4.4.1. A conservation law.

Lemma 4.9. *Let (α, ν) be supercritical data, let $\bar{\varsigma}(\cdot)$ be defined via (28), (29), and (31), and let $\xi_1, \xi_2 \in \mathcal{M}_F^{c,p}$ be such that $\langle \chi, \xi_i\rangle < \infty$ for $i = 1, 2$. If $\xi_1 \preceq \xi_2$, then, for all $t \ge 0$,*

$$\|\bar{M}_1(t,\cdot) - \langle 1_{(\cdot,\infty)}, \bar{\varsigma}(t)\rangle\|_{L_1} = \langle \chi, \xi_1\rangle, \tag{85}$$

$$\|\bar{M}_1(t,\cdot) - \bar{M}_2(t,\cdot)\|_{L_1} = \langle \chi, \xi_2\rangle - \langle \chi, \xi_1\rangle. \tag{86}$$

Proof. By Definition 4.1 and Theorem 4.2, $\langle 1_{(\cdot,\infty)}, \bar{\varsigma}(t)\rangle \le \bar{M}_1(t,\cdot) \le \bar{M}_2(t,\cdot)$ for all $t \ge 0$. Therefore,

$$\begin{aligned}\|\bar{M}_1(t,\cdot) - \bar{M}_2(t,\cdot)\|_{L_1} &= \int_0^\infty (\bar{M}_2(t,x) - \bar{M}_1(t,x))\,dx = \int_0^\infty \bar{M}_2(t,x)\,dx - \int_0^\infty \bar{M}_1(t,x)\,dx\\ &= \langle \chi, \bar{\mu}_2(t)\rangle - \langle \chi, \bar{\mu}_1(t)\rangle = \langle \chi, \xi_2\rangle - \langle \chi, \xi_1\rangle,\end{aligned}$$

where the final equality follows from (23) if the data is strictly supercritical, and from Gromoll et al. [9, Theorem 3.1] if the data is critical. Thus, (86) holds. If the data is strictly supercritical, (85) follows by a similar argument. If the data is critical, (85) is trivial since $\varsigma = \mathbf{0}$ in that case. □

4.4.2. A weak continuity at zero property.

Lemma 4.10. *Let (α, ν) be supercritical data, and let $\bar{\varsigma}(\cdot)$ be defined via (28), (29), and (31). Suppose that $\{\xi_n, n \in \mathbb{N}\} \subset \mathcal{M}_F^{c,p}$ is such that $\xi_n \xrightarrow{w} \mathbf{0}$ and $\langle \chi, \xi_n\rangle \to 0$ as $n \to \infty$. Then, for each $t \ge 0$, $\bar{\mu}_n(t) \xrightarrow{w} \bar{\varsigma}(t)$ as $n \to \infty$.*

Proof. Fix $t \ge 0$. By (85) and the fact that $\langle \chi, \xi_n\rangle \to 0$ as $n \to \infty$, $\lim_{n\to\infty} \|\bar{M}_n(t,\cdot) - \langle 1_{(\cdot,\infty)}, \bar{\varsigma}(t)\rangle\|_{L_1} = 0$. This implies that, for every subsequence $\{n_i\}_{i=1}^\infty$ such that n_i tends to infinity as i tends to infinity, there exists a further subsequence $\{n_{i_j}\}_{j=1}^\infty$ such that, for Lebesgue almost every $x \in \mathbb{R}_+$,

$$\bar{M}_{n_{i_j}}(t,x) \longrightarrow \langle 1_{(x,\infty)}, \bar{\varsigma}(t)\rangle, \quad \text{as } j \to \infty. \tag{87}$$

Notice that, for each $n \in \mathbb{N}$, $\bar{M}_n(t,x)$ is a nonincreasing function of $x \in \mathbb{R}_+$. Also, notice that $\langle 1_{(x,\infty)}, \bar{\varsigma}(t)\rangle$ is continuous for $x \in \mathbb{R}_+$ since $\bar{\varsigma}(t)$ has no atoms. Therefore, the convergence in (87) is in fact pointwise for all $x \in \mathbb{R}_+$, which implies the result. □

4.4.3. The zero initial measure: Uniqueness for strictly supercritical data.

PROOF OF THEOREM 3.2. Let (α, ν) be strictly supercritical data, and let $\bar{\varsigma}(\cdot)$ be defined via (28), (29), and (31). Suppose that $\bar{\mu}(\cdot)$ is a fluid model solution such that $\bar{\mu}(0) = \mathbf{0}$. We must show that $\bar{\mu}(\cdot) = \bar{\varsigma}(\cdot)$. For this, let $\{s_n, n \in \mathbb{N}\} \subset (0, \infty)$ be such that $s_n \to 0$ as $n \to \infty$. For each $n \in \mathbb{N}$, let $\bar{\mu}_n(\cdot) = (\tau_{s_n}\bar{\mu})(\cdot) = \bar{\mu}(\cdot + s_n)$ and $\xi_n = \bar{\mu}_n(0) = \bar{\mu}(s_n)$. By (21), $\xi_n \in \mathcal{M}_F^p$, and by Lemma 3.1, $\bar{\mu}_n(\cdot)$ is a fluid model solution. This, together with Proposition 4.1, implies that $\xi_n \in \mathcal{M}_F^{c,\,p}$. By the continuity of $\bar{\mu}(\cdot)$, $\xi_n \xrightarrow{w} \mathbf{0}$ as $n \to \infty$. In addition, by (23), $\langle \chi, \xi_n \rangle = (\rho - 1)s_n$ for each $n \in \mathbb{N}$. Thus, $\langle \chi, \xi_n \rangle \to 0$ as $n \to \infty$. Therefore, by Lemma 4.10, it follows that for each $t \geq 0$, $\bar{\mu}_n(t) \xrightarrow{w} \bar{\varsigma}(t)$ as $n \to \infty$. But, since $\bar{\mu}(\cdot)$ is continuous, it also follows that $\bar{\mu}_n(t) \xrightarrow{w} \bar{\mu}(t)$ as $n \to \infty$ for each $t \geq 0$. Consequently, $\bar{\mu}(t) = \bar{\varsigma}(t)$ for all $t \geq 0$. □

4.4.4. Additional consequences of the weak continuity property for strictly supercritical data.

LEMMA 4.11. *Let (α, ν) be strictly supercritical data, and let ς be defined by (28) and (29). Any fluid model solution $\bar{\mu}(\cdot)$ for the data (α, ν) that satisfies $\overline{W}(0) < \infty$ also satisfies*

$$\frac{\bar{\mu}(t)}{t} \xrightarrow{\ w\ } \varsigma, \quad \text{as } t \to \infty. \tag{88}$$

PROOF. Let $\bar{\mu}(\cdot)$ be a fluid model solution for the data (α, ν) such that $\langle \chi, \bar{\mu}(0) \rangle < \infty$. If $\bar{\mu}(0) = \mathbf{0}$, then (88) follows immediately from Theorem 3.2. Thus, it is henceforth assumed that $\bar{\mu}(0) \neq \mathbf{0}$. By Lemma 3.1(ii), for each $s > 0$, $(\mathcal{L}_s\bar{\mu})(t) = \bar{\mu}(st)/s$ for all $t \geq 0$ is a fluid model solution for the data (α, ν). By Proposition 4.1 and Remark 4.1, $(\mathcal{L}_s\bar{\mu})(0) \in \mathcal{M}_F^{c,\,p}$ for each $s > 0$. Moreover,

$$(\mathcal{L}_s\bar{\mu})(0) \xrightarrow{\ w\ } \mathbf{0} \qquad \text{and} \qquad \langle \chi, (\mathcal{L}_s\bar{\mu})(0) \rangle \longrightarrow 0, \quad \text{as } s \to \infty.$$

Thus, from Lemma 4.10, it follows that

$$(\mathcal{L}_s\bar{\mu})(1) \xrightarrow{\ w\ } \bar{\varsigma}(1) = \varsigma, \quad \text{as } s \to \infty,$$

and the result follows. □

Similarly, one can prove the following lemma by using Lemma 4.11 in place of Theorem 3.5 in the proof of Theorem 3.6 given in §3.2.7.

LEMMA 4.12. *Let (α, ν) be strictly supercritical data, and let ς be defined by (28) and (29). A fluid model solution $\bar{\mu}(\cdot)$ for the data (α, ν) such that $\overline{W}(0) < \infty$ is stationary if and only if for some $s \in [0, \infty)$, $\bar{\mu}(t) = (\tau_s\bar{\varsigma})(t)$, for all $t \geq 0$.*

4.5. Convergence to $\bar{\varsigma}(\cdot)$. In this section, we use the order preservation property to prove Theorem 3.4. By Theorem 3.2, $\Xi(\mathbf{0}) = \bar{\varsigma}(\cdot)$. Thus, in order to prove Theorem 3.4, it suffices to show that, given $\{\xi_n, n \in \mathbb{N}\} \subseteq \mathcal{M}_F^{c,\,p}$ such that $\xi_n \xrightarrow{w} \mathbf{0}$ as $n \to \infty$, we have, for each $T > 0$ and $g \in \mathbf{C}_b(\mathbb{R}_+)$,

$$\lim_{n\to\infty} \sup_{t\in[0,T]} |\langle g, \bar{\mu}_n(t) \rangle - \langle g, \bar{\varsigma}(t) \rangle| = 0. \tag{89}$$

Recall that for critical data, $\bar{\varsigma}(\cdot) \equiv \mathbf{0}$. Therefore, for critical data, it follows that, for all $g \in \mathbf{C}_b(\mathbb{R}_+)$, $T > 0$, and $n \in \mathbb{N}$,

$$\sup_{t\in[0,T]} |\langle g, \bar{\mu}_n(t) \rangle - \langle g, \bar{\varsigma}(t) \rangle| \leq \|g\|_\infty \sup_{t\in[0,T]} \bar{Z}_n(t). \tag{90}$$

So, in the case of critical data, it suffices to show that the total mass tends to zero uniformly on compact time intervals. In general, uniform control over the convergence of the total mass to a suitable limit is the key ingredient in the proof of (89). In particular, the main step in the proof of Theorem 3.4 is to prove the following theorem.

THEOREM 4.3. *Let (α, ν) be supercritical data, and let $\{\xi_n, n \in \mathbb{N}\} \subset \mathcal{M}_F^{c,\,p}$ be such that $\xi_n \xrightarrow{w} \mathbf{0}$ as $n \to \infty$. Then, for each $T > 0$,*

$$\lim_{n\to\infty} \sup_{t\in[0,T]} |\bar{Z}_n(t) - mt| = 0, \tag{91}$$

where, for strictly supercritical data, m is the unique positive real solution of (3.10), *and for critical data, $m = 0$.*

Recall that $\langle 1, \bar{\varsigma}(t) \rangle = mt$ for all $t \geq 0$ since $\bar{\varsigma}(\cdot) \equiv \mathbf{0}$ for critical data and (30) holds for strictly supercritical data. In particular, for critical data, (89) follows as an immediate consequence of Theorem 4.3 and (90). However, for strictly supercritical data, more extensive analysis is required in order to prove (89) as a consequence of the result in Theorem 4.3. This analysis is contained in §4.5.2. We begin in §4.5.1 by proving Theorem 4.3.

4.5.1. Convergence to $\langle 1, \bar{\varsigma}(\cdot)\rangle$. The proof of Theorem 4.3 depends on the rate at which $U_e(\cdot)$ tends to infinity. For critical data, the elementary renewal theorem implies that $U_e(\cdot)$ tends to infinity linearly. For strictly supercritical data, $U_e(\cdot)$ tends to infinity exponentially fast; this is proved in Lemma 4.13 using the key renewal theorem. We begin this section by reviewing the facts that are needed for the application of this theorem.

Consider strictly supercritical data (α, ν). Our application of the key renewal theorem is confined to a specific renewal function $R(\cdot)$. In order to define $R(\cdot)$, note that, by (27), $\rho\exp(-m\cdot)f_e(\cdot)$ is a probability density function on $\mathbb{R}_+$. For $x\in\mathbb{R}_+$, let

$$C(x) = \rho\int_0^x \exp(-my)f_e(y)\,dy \qquad \text{and} \qquad R(x) = \sum_{i=0}^{\infty} C^{*i}(x). \tag{92}$$

Then, $R(\cdot)$ is the renewal function for a zero-delay renewal process with interarrival distribution determined by the cumulative distribution function $C(\cdot)$. Furthermore, it is easy to verify that $\int_{\mathbb{R}_+}(1-C(x))\,dx<\infty$. The key renewal theorem implies that for any directly Riemann integrable function $g\colon \mathbb{R}_+\to\mathbb{R}_+$,

$$\lim_{z\to\infty}(g*R)(z) = \frac{\int_{\mathbb{R}_+} g(x)\,dx}{\int_{\mathbb{R}_+}(1-C(x))\,dx} \tag{93}$$

(cf. Feller [6, Chapter XI], or Puha and Williams [15, §4] for a concise summary). It is easy to verify that the function $g(x)=\exp(-mx)$, $x\in\mathbb{R}_+$, is directly Riemann integrable since it is Riemann integrable and nonincreasing.

LEMMA 4.13. *Let (α, ν) be strictly supercritical data, and let $C(\cdot)$ and $R(\cdot)$ be defined via (92). Then, for all $t\geq 0$,*

$$U_e(t) = \int_0^t \exp(ms)\,dR(s). \tag{94}$$

In particular,

$$\lim_{t\to\infty}\frac{U_e(t)}{\exp(mt)} = \frac{1}{m\int_0^\infty(1-C(t))\,dt}. \tag{95}$$

PROOF. First, note that (95) is an immediate consequence of (94), the fact that $\exp(-mx)$, $x\in\mathbb{R}_+$, is directly Riemann integrable, and (93). So, it suffices to prove (94). For this, note that for all $t\geq 0$,

$$U_e(t) = 1+\rho(U_e * F_e)(t).$$

For each $t\geq 0$, let

$$\hat{U}_e(t) = \exp(-mt)U_e(t) \qquad \text{and} \qquad \hat{f}_e(t) = \exp(-mt)f_e(t).$$

Then, for each $t\geq 0$,

$$\hat{U}_e(t) = \exp(-mt) + \rho\int_0^t \hat{U}_e(t-s)\hat{f}_e(s)\,ds.$$

In particular, $\hat{U}_e(\cdot)$ satisfies a Volterra equation that is very similar to the Volterra equation (57). So, via the same line of reasoning that was used in the beginning of §4 to argue that (60) is the unique locally bounded, Borel-measureable solution of (57), equation (94) follows (also see Resnick [16, Theorem 3.5.1] or Feller [6, the lemma on p. 359]). □

Next we show how to use Lemma 4.13 to prove Theorem 4.3. This proof is guided by the following intuition: The more time that the initial mass spends in the system, the more opportunity it has to slow down processing and cause the total mass to build up. So the worst-case scenario would be if the initial mass never exited the system, i.e., if the initial measure was an atom at infinity. Clearly, such a measure is not an element of $\mathcal{M}_F$, and therefore cannot be used as an initial measure. However, Proposition 4.2 and Lemma 4.8(ii) allow us to circumvent this technicality. Indeed, given $\xi\in\mathcal{M}_F^{c,p}$, we can choose a positive constant c such that c is strictly larger than the total mass of ξ. Taking a measure that has an atom at infinity of size c corresponds to fixing a function $h(\cdot)$ that is identically equal to c. Then, $h(\cdot)$ is strictly larger than $h_\xi(\cdot)$, and so the conditions in Lemma 4.8(ii) are satisfied. Consequently, the $L(\cdot)$ associated with $h(\cdot)$ dominates the $\bar{Z}_\xi(\cdot)$ associated with ξ. This rigorizes the intuition that throwing all of the mass out to infinity results in a worst-case scenario. A consequence of this is that, in order to prove (91), it suffices to prove the analogue of (91) for a sequence of L_n corresponding to positive, constant h_n which tend to zero and dominate the h_{ξ_n}. The advantage of working with the h_n is that they are constant, and thus the functions determined by the h_n (H_n, J_n, K_n, and L_n) obey certain scaling properties (cf. (97)), as we will see in the proof below.

Proof of Theorem 4.3. By Theorem 4.2, for each $n \in \mathbb{N}$, $mt \le \bar{Z}_n(t)$, for all $t \ge 0$. Fix a monotone, nonincreasing sequence $\{c_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ that converges to zero as $n\to\infty$ and that satisfies $\langle 1, \xi_n\rangle < c_n$ for all $n\in\mathbb{N}$. For each $n\in\mathbb{N}$, let $h_n(\cdot)\equiv c_n$. Then, for each $n\in\mathbb{N}$, since $\langle 1_{(x,\infty)}, \xi_n\rangle < c_n = h_n(x)$ for all $x\in\mathbb{R}_+$, by Lemma 4.8(ii) and Proposition 4.2, we have $\bar{Z}_n(\cdot)\le L_n(\cdot)$. In particular, for each $n\in\mathbb{N}$, $0\le \bar{Z}_n(t)-mt \le L_n(t)-mt$, for all $t\ge 0$. Thus, in order to prove Theorem 4.3, it suffices to show that for each $T>0$,

$$\lim_{n\to\infty} \sup_{t\in[0,T]} (L_n(t)-mt) = 0. \tag{96}$$

In order to prove (96), fix $T>0$ and let $h(x)=1$ for all $x\in\mathbb{R}_+$. Then, for all $t\ge 0$,

$$j(t)=U_e(t), \qquad J(t)=\int_0^t U_e(s)\,ds, \qquad K(t)=J^{-1}(t), \qquad \text{and} \qquad L(t)=j(K(t)).$$

Therefore, since $h_n(\cdot)\equiv c_n$ for each $n\in\mathbb{N}$, it follows that for all $n\in\mathbb{N}$ and $t\ge 0$,

$$j_n(t)=c_n j(t), \qquad J_n(t)=c_n J(t), \qquad K_n(t)=J^{-1}(t/c_n), \qquad \text{and} \qquad L_n(t)=c_n L(t/c_n). \tag{97}$$

Thus, for all $n\in\mathbb{N}$ and any constant $M\in(0,T/c_n]$,

$$\begin{aligned}
\sup_{t\in[0,T]}(L_n(t)-mt) &= \sup_{t\in[0,T]}\left(c_n L\left(\frac{t}{c_n}\right)-mt\right)\\
&\le \sup_{t\in[0,Mc_n]}\left(c_n L\left(\frac{t}{c_n}\right)-mt\right) + \sup_{t\in[Mc_n,T]} t\left(\frac{c_n}{t}L\left(\frac{t}{c_n}\right)-m\right)\\
&\le \sup_{t\in[0,Mc_n]} c_n L\left(\frac{t}{c_n}\right) + \sup_{t\in[Mc_n,T]} t\left(\frac{c_n}{t}L\left(\frac{t}{c_n}\right)-m\right).
\end{aligned}$$

Note that $L(\cdot)$ is nondecreasing since both $j(\cdot)$ and $K(\cdot)$ are nondecreasing. Using this, we obtain that for all $n\in\mathbb{N}$ and any constant $M\in(0,T/c_n]$,

$$\begin{aligned}
\sup_{t\in[0,T]}(L_n(t)-mt) &\le c_n L(M) + T\sup_{t\in[c_nM,T]}\left(\frac{c_n}{t}L\left(\frac{t}{c_n}\right)-m\right)\\
&\le c_n L(M) + T\sup_{t\in[M,T/c_n]}\left(\frac{L(t)}{t}-m\right)\\
&\le c_n L(M) + T\sup_{t\ge M}\left(\frac{L(t)}{t}-m\right).
\end{aligned}$$

Thus, it suffices to show that

$$\lim_{t\to\infty}\frac{L(t)}{t}=m. \tag{98}$$

Indeed, if (98) holds, then, given $\epsilon>0$, there exists a $M\in(0,\infty)$ such that

$$\sup_{t\ge M}\left(\frac{L(t)}{t}-m\right) < \frac{\epsilon}{2T},$$

and an $N\in\mathbb{N}$ such that for all $n\ge N$, we have $T/c_n > M$ and $c_n L(M) < \epsilon/2$. Thus, if (98) holds, then given $\epsilon>0$, there exists an $N\in\mathbb{N}$ such that for all $n\ge N$,

$$\sup_{t\in[0,T]}(L_n(t)-mt) < \epsilon,$$

which proves (96). Therefore, it suffices to prove (98).

In order to prove (98), notice that for all $t\ge 0$, $L(J(t))=j(t)$. Since $J(\cdot)$ is continuous and strictly increasing, and $J(t)$ tends to infinity as t tends to infinity, in order to prove (98), it suffices to show that

$$\lim_{t\to\infty}\frac{j(t)}{J(t)}=m. \tag{99}$$

Since, for $t \geq 0$, $j(t) = U_e(t)$, (99) is equivalent to

$$\lim_{t\to\infty} \frac{U_e(t)}{\int_0^t U_e(s)\,ds} = m. \tag{100}$$

So, in order to prove (98), it suffices to show that (100) holds.

To prove (100), there are two cases to consider. First, suppose that the data is critical. Then $m = 0$, and by the elementary renewal theorem, $U_e(t)/t$ converges to $\beta_e = 1/\langle \chi, \nu_e\rangle$ as t tends to infinity, where $\beta_e = 0$ if $\langle \chi, \nu_e\rangle = \infty$ (cf. Resnick [16, Theorem 3.3.3]). If $\beta_e > 0$, then (100) follows immediately. Otherwise, $\beta_e = 0$. In this case, since $U_e(t) \geq 1$ for all $t \geq 0$, it follows that

$$\frac{U_e(t)}{\int_0^t U_e(s)\,ds} \leq \frac{U_e(t)}{t}, \quad \text{for all } t \geq 0,$$

and (100) follows. Therefore, (100) holds for critical data. For strictly supercritical data, by Lemma 4.13 and L'Hôpital's rule, there exists a $c \in (0, \infty)$ such that

$$\lim_{t\to\infty} \frac{U_e(t)}{\exp(mt)} = c \qquad \text{and} \qquad \lim_{t\to\infty} \frac{\int_0^t U_e(s)\,ds}{\exp(mt)} = \frac{c}{m}. \tag{101}$$

Therefore,

$$\lim_{t\to\infty} \frac{U_e(t)}{\int_0^t U_e(s)\,ds} = m.$$

So (100) holds for strictly supercritical data as well. □

4.5.2. Proof of Theorem 3.4. In the case of critical data, Theorem 3.4 is an immediate consequence of Theorem 4.3. So the main focus here is to show how to use Theorem 4.3 in the case of strictly supercritical data to prove Theorem 3.4. The first order of business is to obtain a bound that serves the same purpose for strictly supercritical data as (90) serves for critical data. For this, we derive analogues of (61) and (68) with $g \in \mathbf{C}_b(\mathbb{R}_+)$ in place of the indicator functions. These analogues involve shifted versions of the g of the form $g(\cdot - a)$, for $a \geq 0$. For this, recall that we have adopted the convention that all g are extended to be identically equal to zero on the negative half line.

Lemma 4.14. *Let (α, ν) be supercritical data, and let $\xi \in \mathscr{M}_F^{c,p}$. For all $t \geq 0$ and $g \in \mathbf{C}_b(\mathbb{R}_+)$,*

$$\langle g, \bar{\mu}_\xi(t)\rangle = \langle g(\cdot - \bar{S}_\xi(t)), \xi\rangle + \alpha \int_0^t \langle g(\cdot - (\bar{S}_\xi(t) - \bar{S}_\xi(s))), \nu\rangle\,ds. \tag{102}$$

In addition, if (α, ν) is strictly supercritical data and $\bar{\varsigma}(\cdot)$ is defined via (28), (29), and (31), then for all $t > 0$ and $g \in \mathbf{C}_b(\mathbb{R}_+)$,

$$\langle g, \bar{\varsigma}(t)\rangle = \alpha \int_{(0,t]} \left\langle g\left(\cdot - \frac{1}{m}\log\left(\frac{t}{s}\right)\right), \nu\right\rangle ds. \tag{103}$$

Proof. Fix supercritical data (α, ν) and $\xi \in \mathscr{M}_F^{c,p}$. To prove (102) and (103), we use a monotone class theorem for functions (cf. Durrett [4, Chapter 5, Theorem 1.4]). Since the two arguments are very similar, we prove only (102) here and leave the proof of (103) for the reader. To begin, let

$$\mathscr{H} = \{g\colon \mathbb{R}_+ \to \mathbb{R}\colon g \text{ is Borel measurable, bounded and satisfies (102)}\}.$$

We wish to show that $\mathbf{C}_b(\mathbb{R}_+) \subset \mathscr{H}$. Let

$$\mathscr{A} = \{[x, w)\colon 0 \leq x < w < \infty\} \cup \varnothing.$$

Then, $\mathscr{A}$ is a π-system that generates the Borel σ-algebra on $\mathbb{R}_+$. Also, note that for all $t \geq 0$ and $0 \leq x < w \leq \infty$,

$$g = 1_{[x,w)} \in \mathscr{H}.$$

For $x \neq 0$, the above follows from (61) if the data is strictly supercritical, and from Gromoll et al. [9, Lemma 4.3] if the data is critical. For strictly supercritical data (resp. critical data), to show that the above holds for $x = 0$, use (61) (resp. [9, Lemma 4.3]), (S.2) (resp. (C.2)), and the following facts: ξ has no atoms, $\bar{S}_\xi(\cdot)$

is strictly increasing, and ν has at most countably many atoms. Thus, $A \in \mathcal{A}$ implies that $1_A \in \mathcal{H}$. It is easily verified that $f, g \in \mathcal{H}$ implies that $f + g \in \mathcal{H}$ and $cf \in \mathcal{H}$ for all $c \in \mathbb{R}$. Finally, fix $\{g_n, n \in \mathbb{N}\} \subset \mathcal{H}$ such that for each $n \in \mathbb{N}$, g_n is nonnegative and the sequence $\{g_n, n \in \mathbb{N}\}$ increases to a bounded, Borel-measurable function $g: \mathbb{R}_+ \to \mathbb{R}$. To complete the proof, it suffices to show that $g \in \mathcal{H}$. By the monotone convergence theorem, for each $0 \le s \le t$,

$$\lim_{n\to\infty} \langle g_n(\cdot), \bar{\mu}_\xi(t)\rangle = \langle g(\cdot), \bar{\mu}_\xi(t)\rangle,$$

$$\lim_{n\to\infty} \langle g_n(\cdot - \bar{S}_\xi(t)), \xi\rangle = \langle g(\cdot - \bar{S}_\xi(t)), \xi\rangle,$$

and

$$\lim_{n\to\infty} \langle g_n(\cdot - (\bar{S}_\xi(t) - \bar{S}_\xi(s))), \nu\rangle = \langle g(\cdot - (\bar{S}_\xi(t) - \bar{S}_\xi(s))), \nu\rangle.$$

This, together with $g_n \in \mathcal{H}$ for each $n \in \mathbb{N}$, implies that $g \in \mathcal{H}$ and this completes the proof of (102). □

For $\{\xi_n, n \in \mathbb{N}\} \subset \mathcal{M}_F^{c,p}$ such that $\xi_n \xrightarrow{w} \mathbf{0}$ as $n \to \infty$, from (102) and (103) it follows that, if the data is strictly supercritical, then for all $g \in \mathbf{C}_b(\mathbb{R}_+)$, $T > 0$, and $n \in \mathbb{N}$,

$$\sup_{t\in[0,T]} |\langle g, \bar{\mu}_n(t)\rangle - \langle g, \bar{\varsigma}(t)\rangle|$$
$$\le \|g\|_\infty \langle 1, \xi_n\rangle + \alpha \sup_{t\in[0,T]} \int_{(0,t]} \left\langle \left| g(\cdot - (\bar{S}_n(t) - \bar{S}_n(s))) - g\left(\cdot - \frac{1}{m}\log\left(\frac{t}{s}\right)\right)\right|, \nu\right\rangle ds. \tag{104}$$

Since g is bounded and continuous, what is needed is some control over the convergence of $\bar{S}_n(t) - \bar{S}_n(s)$ to $(1/m)\log(t/s)$ as $n \to \infty$, for each $0 < s \le t < \infty$, which can be obtained from Theorem 4.3.

Lemma 4.15. *Let (α, ν) be strictly supercritical data, and let $\{\xi_n, n \in \mathbb{N}\} \subset \mathcal{M}_F^{c,p}$ be such that $\xi_n \xrightarrow{w} \mathbf{0}$ as $n \to \infty$. Then, for each $0 < T_1 \le T_2 < \infty$,*

$$\lim_{n\to\infty} \sup_{T_1 \le s\le t\le T_2} \left| \bar{S}_n(t) - \bar{S}_n(s) - \frac{1}{m}\log\left(\frac{t}{s}\right)\right| = 0.$$

Proof. Fix $0 < T_1 \le T_2 < \infty$. Recall that, by (21), $\bar{Z}_n(t) > 0$ for all $t \in [T_1, T_2]$. Thus, for each $T_1 \le s \le t \le T_2$,

$$\left| \bar{S}_n(t) - \bar{S}_n(s) - \frac{1}{m}\log\left(\frac{t}{s}\right)\right| \le \int_s^t \left| \frac{1}{\bar{Z}_n(u)} - \frac{1}{mu}\right| du.$$

Therefore,

$$\sup_{T_1\le s\le t\le T_2} \left| \bar{S}_n(t) - \bar{S}_n(s) - \frac{1}{m}\log\left(\frac{t}{s}\right)\right| \le (T_2 - T_1) \sup_{t\in[T_1,T_2]} \left| \frac{1}{\bar{Z}_n(t)} - \frac{1}{mt}\right|. \tag{105}$$

From Theorem 4.3, since $T_1 > 0$,

$$\lim_{n\to\infty} \sup_{t\in[T_1,T_2]} \left| \frac{1}{\bar{Z}_n(t)} - \frac{1}{mt}\right| = 0. \tag{106}$$

The result follows by letting $n \to \infty$ in (105) and using (106). □

Proof of Theorem 3.4. First, consider the case where the data is critical. Then, the result in Theorem 3.4 follows from (89), (90), and Theorem 4.3.

Next, consider the case where the data is strictly supercritical. For this, fix strictly supercritical data (α, ν). By (104), it suffices to show that for each $T > 0$, $g \in \mathbf{C}_b(\mathbb{R}_+)$, and $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that, for all $n \ge N$,

$$\sup_{t\in[0,T]} \int_{(0,t]} \left\langle \left| g(\cdot - (\bar{S}_n(t) - \bar{S}_n(s))) - g\left(\cdot - \frac{1}{m}\log\left(\frac{t}{s}\right)\right)\right|, \nu\right\rangle ds \le \epsilon. \tag{107}$$

For this, fix $g \in \mathbf{C}_b(\mathbb{R}_+)$ and $\epsilon > 0$. Let $T_1 = \epsilon/8(\|g\|_\infty + 1)$. It suffices to prove (107) for $T > T_1$, so fix $T > T_1$. Since ν is a probability measure, we obtain, for all $t \in [0, T_1]$ and $n \in \mathbb{N}$,

$$\int_{(0,t]} \left\langle \left| g(\cdot - (\bar{S}_n(t) - \bar{S}_n(s))) - g\left(\cdot - \frac{1}{m}\log\left(\frac{t}{s}\right)\right)\right|, \nu\right\rangle ds \le 2\|g\|_\infty t \le \frac{\epsilon}{4}. \tag{108}$$

Next, consider $t \in [T_1, T]$. For each $n \in \mathbb{N}$,

$$\int_{(0,t]} \left\langle \left| g(\cdot - (\bar{S}_n(t) - \bar{S}_n(s))) - g\left(\cdot - \frac{1}{m}\log\left(\frac{t}{s}\right)\right)\right|, \nu \right\rangle ds$$
$$\leq \frac{\epsilon}{4} + \int_{T_1}^{t} \left\langle \left| g(\cdot - (\bar{S}_n(t) - \bar{S}_n(s))) - g\left(\cdot - \frac{1}{m}\log\left(\frac{t}{s}\right)\right)\right|, \nu \right\rangle ds. \tag{109}$$

Thus, for each $t \in [T_1, T]$, we need to obtain estimates on the integrand of the second term on the right side of (109), for each $s \in [T_1, t]$. In order to obtain such an estimate, let $M \geq 0$ be such that

$$\langle 1_{[M,\infty)}, \nu \rangle \leq \frac{T_1}{T - T_1} = \frac{\epsilon}{8(\|g\|_\infty + 1)(T - T_1)}. \tag{110}$$

By Theorem 4.2, $mt \leq \bar{Z}_n(t)$ for all $t \geq 0$ and $n \in \mathbb{N}$. Therefore, for all $T_1 \leq s \leq t \leq T$ and $n \in \mathbb{N}$, $\bar{S}_n(t) - \bar{S}_n(s) \leq (1/m)\log(t/s)$. Thus, for all $T_1 \leq s \leq t \leq T$ and $n \in \mathbb{N}$, using the fact the $g(\cdot)$ is zero on the negative half line, we have

$$\left\langle \left| g(\cdot - (\bar{S}_n(t) - \bar{S}_n(s))) - g\left(\cdot - \frac{1}{m}\log\left(\frac{t}{s}\right)\right)\right|, \nu \right\rangle$$
$$\leq \int_{[\bar{S}_n(t) - \bar{S}_n(s),\, (1/m)\log(t/s))} |g(x - (\bar{S}_n(t) - \bar{S}_n(s)))|\, \nu(dx)$$
$$+ \int_{[(1/m)\log(t/s),\, M)} \left| g(x - (\bar{S}_n(t) - \bar{S}_n(s))) - g\left(x - \frac{1}{m}\log\left(\frac{t}{s}\right)\right)\right| \nu(dx)$$
$$+ \int_{[M,\infty)} \left| g(x - (\bar{S}_n(t) - \bar{S}_n(s))) - g\left(x - \frac{1}{m}\log\left(\frac{t}{s}\right)\right)\right| \nu(dx)$$
$$\leq \|g\|_\infty \langle 1_{[\bar{S}_n(t) - \bar{S}_n(s),\, (1/m)\log(t/s))}, \nu \rangle$$
$$+ \int_{[(1/m)\log(t/s),\, M)} \left| g(x - (\bar{S}_n(t) - \bar{S}_n(s))) - g\left(x - \frac{1}{m}\log\left(\frac{t}{s}\right)\right)\right| \nu(dx)$$
$$+ 2\|g\|_\infty \langle 1_{[M,\infty)}, \nu \rangle. \tag{111}$$

Let $\eta > 0$ be such that

$$|g(x) - g(y)| \leq \frac{\epsilon}{4(T - T_1)}, \quad \text{for all } x, y \in [0, M] \text{ such that } |x - y| < \eta.$$

For each $\eta' \in (0, \eta]$, let $N(\eta') \in \mathbb{N}$ be such that

$$\sup_{T_1 \leq s \leq t \leq T} \left| \bar{S}_n(t) - \bar{S}_n(s) - \frac{1}{m}\log\left(\frac{t}{s}\right)\right| < \eta', \quad \text{for all } n \geq N(\eta').$$

(Note that such an $N(\eta')$ exists by Lemma 4.15.) Observing that the arguments of $g(\cdot)$ in the second to the last term in (111) all lie in $[0, M]$, for $T_1 \leq s \leq t \leq T$ and $\eta' \in (0, \eta]$ and using (110) on the last term in (111), we have for all $n \geq N(\eta')$,

$$\left\langle \left| g(\cdot - (\bar{S}_n(t) - \bar{S}_n(s))) - g\left(\cdot - \frac{1}{m}\log\left(\frac{t}{s}\right)\right)\right|, \nu \right\rangle$$
$$\leq \|g\|_\infty \langle 1_{[\bar{S}_n(t) - \bar{S}_n(s),\, (1/m)\log(t/s))}, \nu \rangle + \frac{\epsilon}{4(T - T_1)} \langle 1_{[(1/m)\log(t/s),\, M)}, \nu \rangle + \frac{\epsilon}{4(T - T_1)}$$
$$\leq \|g\|_\infty \langle 1_{[((1/m)\log(t/s) - \eta')^+,\, (1/m)\log(t/s))}, \nu \rangle + \frac{\epsilon}{4(T - T_1)} + \frac{\epsilon}{4(T - T_1)}.$$

Therefore, for $t \in [T_1, T]$ and $\eta' \in (0, \eta]$, we have for all $n \geq N(\eta')$,

$$\int_{T_1}^{t} \left\langle \left| g(\cdot - (\bar{S}_n(t) - \bar{S}_n(s))) - g\left(\cdot - \frac{1}{m}\log\left(\frac{t}{s}\right)\right)\right|, \nu \right\rangle ds \leq \frac{\epsilon}{2} + \|g\|_\infty \int_{T_1}^{t} \langle 1_{[((1/m)\log(t/s) - \eta')^+,\, (1/m)\log(t/s))}, \nu \rangle\, ds. \tag{112}$$

Using (109), the definition of T_1, and (112), we obtain the following. For all $t \in [T_1, T]$ and $\eta' \in (0, \eta]$, we have for all $n \geq N(\eta')$,

$$\int_{(0,t]} \left\langle \left| g(\cdot - (\bar{S}_n(t) - \bar{S}_n(s))) - g\left(\cdot - \frac{1}{m}\log\left(\frac{t}{s}\right)\right)\right|, \nu \right\rangle ds \leq \frac{3\epsilon}{4} + \|g\|_\infty \int_{T_1}^t \langle 1_{[((1/m)\log(t/s)-\eta')^+,\,(1/m)\log(t/s))}, \nu\rangle\, ds. \tag{113}$$

So, in order to complete the proof of (107), it suffices to show that there exists $\eta' \in (0, \eta]$ such that for all $t \in [T_1, T]$,

$$\int_{T_1}^t \langle 1_{[((1/m)\log(t/s)-\eta')^+,\,(1/m)\log(t/s))}, \nu\rangle\, ds \leq \frac{\epsilon}{4(\|g\|_\infty + 1)}. \tag{114}$$

This requires some care since ν can have atoms. Let $A \subset \mathbb{R}_+$ denote the set containing all of the atoms of ν, which is at most countable, and let $\nu_d = \sum_{a\in A} \nu(\{a\})\delta_a$ be the Borel measure comprised of only the atoms of ν. Then, $\nu_c = \nu - \nu_d$ has no atoms. Therefore, there exists $\eta_1' \in (0, \eta]$ such that $\sup_{y\in\mathbb{R}_+} \langle 1_{(y-\eta_1',\, y+\eta_1')}, \nu_c\rangle < \epsilon/(12(T-T_1)(\|g\|_\infty + 1))$ (cf. the proof of Gromoll et al. [9, Lemma A.1]). Thus, for $t \in [T_1, T]$ and $\eta' \in (0, \eta_1']$, we have

$$\begin{aligned} &\int_{T_1}^t \langle 1_{[((1/m)\log(t/s)-\eta')^+,\,(1/m)\log(t/s))}, \nu\rangle\, ds \\ &\quad= \int_{T_1}^t \langle 1_{[((1/m)\log(t/s)-\eta')^+,\,(1/m)\log(t/s))}, \nu_c\rangle\, ds + \int_{T_1}^t \langle 1_{[((1/m)\log(t/s)-\eta')^+,\,(1/m)\log(t/s))}, \nu_d\rangle\, ds \\ &\quad\leq \frac{\epsilon}{12(\|g\|_\infty + 1)} + \sum_{a\in A} \nu(\{a\}) \int_{T_1}^t \langle 1_{[((1/m)\log(t/s)-\eta')^+,\,(1/m)\log(t/s))}, \delta_a\rangle\, ds. \end{aligned}$$

Since $\sum_{a\in A} \nu(\{a\}) \leq 1$, there exists a finite set $A_\epsilon \subset A$ such that

$$\sum_{a\in A\setminus A_\epsilon} \nu(\{a\}) \leq \frac{\epsilon}{12(T-T_1)(\|g\|_\infty + 1)}.$$

Thus, for $t \in [T_1, T]$ and $\eta' \in (0, \eta_1']$, we have

$$\begin{aligned} &\int_{T_1}^t \langle 1_{[((1/m)\log(t/s)-\eta')^+,\,(1/m)\log(t/s))}, \nu\rangle\, ds \\ &\quad\leq \frac{\epsilon}{12(\|g\|_\infty + 1)} + \sum_{a\in A\setminus A_\epsilon} \nu(\{a\})(T-T_1) + \sum_{a\in A_\epsilon} \nu(\{a\}) \int_{T_1}^t \langle 1_{[((1/m)\log(t/s)-\eta')^+,\,(1/m)\log(t/s))}, \delta_a\rangle\, ds \\ &\quad\leq \frac{\epsilon}{6(\|g\|_\infty + 1)} + \sum_{a\in A_\epsilon} \nu(\{a\}) \int_{T_1}^t \langle 1_{[((1/m)\log(t/s)-\eta')^+,\,(1/m)\log(t/s))}, \delta_a\rangle\, ds. \end{aligned}$$

If $A_\epsilon = \varnothing$, let $\eta' = \eta_1'$. Otherwise, $A_\epsilon \neq \varnothing$. In that case, for $\eta' \in (0, \eta_1']$, $T_1 \leq s \leq t \leq T$, and $a \in A_\epsilon$, we have

$$a \in \left[\left(\frac{1}{m}\log\left(\frac{t}{s}\right) - \eta'\right)^+, \frac{1}{m}\log\left(\frac{t}{s}\right)\right) \quad \text{only if } s \in [t\exp(-m(a+\eta')),\, t\exp(-ma)).$$

Therefore, for each $\eta' \in (0, \eta_1']$, $t \in [T_1, T]$, and $a \in A_\epsilon$,

$$\int_{T_1}^t \langle 1_{[((1/m)\log(t/s)-\eta')^+,\,(1/m)\log(t/s))}, \delta_a\rangle\, ds \leq t(\exp(-ma) - \exp(-m(a+\eta'))).$$

So, if $A_\epsilon \neq \varnothing$, let $\eta' \in (0, \eta_1']$ be such that

$$\max_{a\in A_\epsilon}(\exp(-ma) - \exp(-m(a+\eta'))) \leq \frac{\epsilon}{12|A_\epsilon|(\|g\|_\infty + 1)T},$$

where $|A_\epsilon|$ denotes the number of atoms in A_ϵ. Then, on combining the above inequalities, we have, for $t \in [T_1, T]$, verified (114). Combining (108), (113), and (114) proves (107) with $N = N(\eta')$. □

5. Proof of the fluid limit result for strictly supercritical data. In this section, we prove Theorem 3.7. As we will see, if the total mass of the limiting initial measure is bounded away from zero, then the proof is a simple extension of the proof of Gromoll et al. [9, Theorem 3.2]. However, to handle the case where the total mass of the limiting initial measure is not bounded away from zero, new analysis is required. Even so, the overall approach is basically the same as that in [9]. In particular, in order to prove Theorem 3.7, the following theorem is the key result.

THEOREM 5.1. *Let (α, ν) be strictly supercritical data. Consider a sequence of overloaded processor sharing queues as defined in §3.3.1, satisfying assumptions (45)–(52) for the data (α, ν). Then, the sequence of fluid-scaled measure-valued processes $\{\bar{\mu}^r(\cdot)\}$ is tight. Moreover, any limit point $\bar{\mu}^\star(\cdot)$ is a.s. a fluid model solution for the strictly supercritical data (α, ν).*

Recall that the sequence $\{\bar{\mu}^r(\cdot)\}_{r>0}$ is tight if and only if the associated sequence of probability laws on $D([0,\infty), \mathcal{M}_F)$ is tight. The term "limit point" in the above statement refers to any limit in distribution along some subsequence of $\{\bar{\mu}^r(\cdot)\}_{r>0}$. We use this terminology because our objective is to show that all such limit points have the same law. This uniqueness in law will hinge on the almost sure characterization of limit points as fluid model solutions.

Throughout this section, we assume that the conditions in §3.3.1 hold. This section is organized as follows. In §5.1, the dynamic equation satisfied by $\bar{\mu}^r(\cdot)$ for each $r>0$ is given. This dynamic equation is used both in the proof of tightness and the almost sure characterization of limit points. Tightness of $\{\bar{\mu}^r(\cdot)\}_{r>0}$ is proved in §5.2, and the almost sure characterization of limit points as fluid model solutions is proved in §5.3. Finally, the proof of Theorem 3.7 appears in §5.4.

5.1. A dynamic equation. Recall the convention that any function g that is defined on $\mathbb{R}_+$ is extended to be identically equal to zero on $(-\infty, 0)$ so that for all $a>0$, $g(\cdot - a)$ is well defined on $\mathbb{R}_+$.

LEMMA 5.1. *For each $r>0$, almost surely, for each bounded, Borel-measurable function $g\colon \mathbb{R}_+ \to \mathbb{R}$, and for all $t, h \geq 0$,*

$$\langle g, \bar{\mu}^r(t+h)\rangle = \langle (1_{(0,\infty)}g)(\cdot - \bar{S}^r_{t,t+h}), \bar{\mu}^r(t)\rangle + \frac{1}{r}\sum_{i=r\bar{E}^r(t)+1}^{r\bar{E}^r(t+h)} (1_{(0,\infty)}g)(v_i^r - \bar{S}^r_{U_i^r/r,\, t+h}). \tag{115}$$

PROOF. To verify that (115) holds, fix $r>0$, a bounded, Borel-measurable function $g\colon \mathbb{R}_+ \to \mathbb{R}$, and $t, h \geq 0$. Using (13), subtract the equation for $\langle (1_{(0,\infty)}g)(\cdot - \bar{S}^r_{t,t+h}), \mu^r(rt)\rangle$ from the equation for $\langle g, \mu^r(rt+rh)\rangle$. Then, use the definition of the residual service times to cancel the common terms. Finally, divide the result by r to obtain (115). ☐

Equation (115) is referred to as the *dynamic equation*.

5.2. Proof of tightness. In this section, we prove the first half of Theorem 5.1, i.e., that the sequence of measure-valued processes $\{\bar{\mu}^r(\cdot)\}_{r>0}$ is tight. Following the approach in Gromoll et al. [9], to verify tightness of $\{\bar{\mu}^r(\cdot)\}$ we show the following two properties.

(T.1) For each $T>0$ and $0<\eta<1$, there is a compact subset $C_{T,\eta}$ of $\mathcal{M}_F$ such that

$$\liminf_{r\to\infty} \mathbf{P}(\bar{\mu}^r(t) \in C_{T,\eta} \text{ for all } t \in [0,T]) \geq 1-\eta.$$

(T.2) For each $g \in \mathbf{C}_b^1(\mathbb{R}_+)$, the sequence of real-valued processes

$$\{\langle g, \bar{\mu}^r(\cdot)\rangle\}_{r>0} \text{ is tight.}$$

To prove (T.2), we verify that the usual compact containment and controlled-oscillations conditions are satisfied by $\{\langle g, \bar{\mu}^r(\cdot)\rangle\}_{r>0}$ for each $g \in \mathbf{C}_b^1(\mathbb{R}_+)$ (cf. (146) and (147) in Theorem 5.3 below). Once these conditions are established, we proceed to verify (T.1), and thereby complete the proof of tightness (cf. Proof of Tightness for Theorem 5.1 at the end of this subsection). Verifying the controlled-oscillations condition (147) for $\{\langle g, \bar{\mu}^r(\cdot)\rangle\}_{r>0}$, for each $g \in \mathbf{C}_b^1(\mathbb{R}_+)$ is the major challenge in proving Theorem 5.3. In the next few paragraphs, we outline our approach to proving this.

In Gromoll et al. [9], the controlled-oscillations result was proved by considering two cases, namely, the case where in fluid scale the initial workload is small, and the case where it is not small. Note that for heavily loaded processor sharing queues, with high probability, in fluid scale the workload is essentially constant over compact time intervals. This fact is used in [9] to show that if the initial workload is small in fluid scale, then with high probability the initial queue length is small in fluid scale, and it remains small over compact time intervals. Note that if the queue length remains small in fluid scale, for each $g \in \mathbf{C}_b^1(\mathbb{R}_+)$ and $r > 0$, $\langle g, \bar{\mu}^r(\cdot)\rangle$ can only have small oscillations. However, for overloaded processor sharing queues, in fluid scale the workload and thus the queue length does not remain small. Indeed, they grow. Therefore, the proof of the controlled-oscillations condition in Gromoll et al. [9] doesn't immediately carry over to the present setting. Nevertheless, many of the ideas in [9] for proving tightness of $\{\langle g, \bar{\mu}^r(\cdot)\rangle\}_{r>0}$, for $g \in \mathbf{C}_b^1(\mathbb{R}_+)$, are still useful here. Specifically, it turns out that if the initial workload in fluid scale is not small, then the proof of the controlled-oscillations condition extends without change to the present setting. Indeed, Theorem 5.2 below gives sufficient conditions for the controlled-oscillations condition for $\{\langle g, \bar{\mu}^r(\cdot)\rangle\}_{r>0}$, $g \in \mathbf{C}_b^1(\mathbb{R}_+)$, to hold on a compact time interval $[s, T]$; these conditions include the assumptions that the fluid-scaled state descriptor at time s has no large atoms, and that the fluid-scaled workload is bounded below on $[s, T]$. The proof of Theorem 5.2 for $s = 0$ is essentially contained in [9]; however, such a theorem is not explicitly stated there. Here, the main ideas of the proofs in [9] that imply an $s = 0$ version of Theorem 5.2 are reorganized and presented to give a proof that also includes $s > 0$.

In order to identify conditions under which the sufficient conditions of Theorem 5.2 are satisfied, following Gromoll et al. [9] we define a "good" event B^r, $r > 0$, that has probability tending to one as $r \to \infty$ and on which the fluid-scaled primitive processes, initial condition, and workload process are close to their respective fluid limits (cf. Lemma 5.2). To bound the fluid-scaled workload from below, we introduce a second event D_γ^r on which the fluid-scaled initial workload is small. On $B^r \cap \breve{D}_\gamma^r$, where $\breve{D}_\gamma^r$ denotes the complement of D_γ^r, it turns out that the fluid-scaled workload is bounded below from time zero on (cf. Lemma 5.5). In particular, the sufficient conditions of Theorem 5.2 hold on $B^r \cap \breve{D}_\gamma^r$ for $s = 0$. On $B^r \cap D_\gamma^r$, we show that there is a positive, deterministic time T_ϵ, after which the fluid-scaled workload is bounded below (cf. Lemma 5.5). In order to verify that the sufficient conditions of Theorem 5.2 hold on $B^r \cap D_\gamma^r$ for $s = T_\epsilon$, there is still a need to verify that no large atoms have formed by time T_ϵ. For this, we will show that, despite the fact that the fluid-scaled queue length tends to grow, on $B^r \cap D_\gamma^r$ it does so in a controlled manner. Thus, on $B^r \cap D_\gamma^r$, the fluid-scaled queue length remains relatively small up to the time T_ϵ (cf. Lemma 5.6). In particular, no large atoms form by time T_ϵ, since there simply isn't much mass in the system by the time T_ϵ.

On $B^r \cap D_\gamma^r$, there is also a need to prove the controlled-oscillations condition up to the time T_ϵ. In fact, it is actually necessary to maintain some sort of control for a time slightly beyond T_ϵ in order to rule out big oscillations at T_ϵ. This is easily handled since, on $B^r \cap D_\gamma^r$, the fluid-scaled queue length is small (even up to time $2T_\epsilon$), and therefore does not admit large oscillations (cf. the proof of Theorem 5.3).

We begin by stating a preliminary lemma in which we define the so-called good events B^r, for $r > 0$, and give a lower bound on the probability of B^r for r sufficiently large.

LEMMA 5.2. *Let (α, ν) be strictly supercritical data. Consider a sequence of processor sharing queues as defined in §3.3.1, satisfying assumptions (45)–(52) for data (α, ν). Let $T > (4\alpha)^{-1}$ and $0 < \epsilon$, $\eta < 1$ be given. There exist strictly positive constants T_ϵ, l, M_0, M_T, γ, K, Γ, κ, and r_0, and events $\{B^r\}_{r>0}$, such that if $r > r_0$, then $\mathbf{P}(B^r) \geq 1 - \eta$, and on B^r, the following hold for all $r > 0$:*

$$T_\epsilon = \frac{\epsilon}{8\alpha} \tag{116}$$

$$l < T_\epsilon \tag{117}$$

$$\sup_{t \in [0, T-l]} \bar{E}^r(t + l) - \bar{E}^r(t) < \frac{\epsilon}{4} \tag{118}$$

$$\bar{E}^r(t) < 2\alpha t, \quad t \in [0, T] \tag{119}$$

$$\langle \chi, \bar{\mu}^r(0)\rangle \vee \langle 1, \bar{\mu}^r(0)\rangle < M_0 \tag{120}$$

$$\langle 1_{[M_0, \infty)}, \Theta\rangle < \frac{\epsilon}{4} \tag{121}$$

$$M_T = M_0 + 2\alpha T \tag{122}$$

$$\kappa < \frac{l}{2M_T} \tag{123}$$

$$\sup_{x \in \mathbb{R}_+} \langle 1_{[x, x+\kappa]}, \bar{\mu}^r(0)\rangle < \frac{\epsilon}{8}, \tag{124}$$

$$\gamma < \min\left\{\frac{\kappa\epsilon}{4}, 2T_\epsilon(\rho - 1)\right\} \tag{125}$$

$$\sup_{t\in[0,T]} |\langle \chi, \bar{\mu}^r(t)\rangle - \langle \chi, \bar{\mu}^r(0)\rangle - (\rho-1)t| < \frac{\gamma}{4} \tag{126}$$

$$\sup_{t\in[0,T]} \frac{1}{r}\sum_{i=1}^{r\bar{E}^r(t)} v_i^r 1_{\{v_i^r>K\}} + \langle \chi 1_{(K,\infty)}, \bar{\mu}^r(0)\rangle < \frac{\gamma}{5} \tag{127}$$

$$\Gamma = K\left(\frac{\gamma}{4}-\frac{\gamma}{5}\right)^{-1}, \tag{128}$$

and for all $n\in\{0,1,\ldots,N\}$, *where* $N=\lceil T\Gamma/\kappa\rceil$,

$$\sup_{t\in[0,T-l]} \frac{1}{r}\sum_{i=r\bar{E}^r(t)+1}^{r\bar{E}^r(t+l)} 1_{[n\kappa,(n+1)\kappa)}(v_i^r) < \frac{\epsilon}{4}\langle 1_{[(n-1/2)\kappa,(n+3/2)\kappa)}, \nu\rangle. \tag{129}$$

The proof of Lemma 5.2 is omitted as it is much like the proof of Gromoll et al. [9, Lemma 5.2]; however, we note that a key ingredient is a functional law of large numbers for E^r and V^r. We now summarize some immediate consequences. First,

$$2T_\epsilon \le T \qquad \text{and} \qquad T_\epsilon < T-l. \tag{130}$$

By (115), for $t\ge 0$,

$$\langle 1, \bar{\mu}^r(t)\rangle \le \langle 1, \bar{\mu}^r(0)\rangle + \bar{E}^r(t). \tag{131}$$

This, together with (119), (120), and (122), implies that on B^r,

$$\sup_{t\in[0,T]} \langle 1, \bar{\mu}^r(t)\rangle \le M_T. \tag{132}$$

In addition, by (126), on B^r,

$$\langle \chi, \bar{\mu}^r(0)\rangle + (\rho-1)t - \frac{\gamma}{4} \le \langle \chi, \bar{\mu}^r(t)\rangle \le \langle \chi, \bar{\mu}^r(0)\rangle + (\rho-1)t + \frac{\gamma}{4}, \quad t\in[0,T]. \tag{133}$$

The next theorem gives sufficient conditions under which the controlled-oscillations condition holds.

THEOREM 5.2. *Let* $g\in \mathbf{C}_b^1(\mathbb{R}_+)$, $T>(4\alpha)^{-1}$, *and* $0<\tilde{\epsilon},\eta<1$. *Set* $\epsilon=\tilde{\epsilon}/2(\|g\|_\infty\vee 1)$. *Let* T_ϵ, l, M_0, M_T, γ, K, Γ, κ, *and* r_0 *be the constants, and* $\{B^r\}_{r>0}$ *be the events, given by Lemma 5.2. Set*

$$\delta = \min\left\{l, \frac{\tilde{\epsilon}}{4\Gamma M_T(\|g'\|_\infty\vee 1)}, \frac{\kappa}{\Gamma}, 1\right\}. \tag{134}$$

Suppose that $0\le s<T-l$, $0<r$, *and* C_s^r *is an event such that* $C_s^r\subset B^r$, *and on* C_s^r,

$$\sup_{x\in\mathbb{R}_+} \langle 1_{[x,x+\kappa]}, \bar{\mu}^r(s)\rangle \le \frac{\epsilon}{2} \tag{135}$$

and

$$\inf_{t\in[s,T]} \langle \chi, \bar{\mu}^r(t)\rangle \ge \frac{\gamma}{4}. \tag{136}$$

Then, on C_s^r,

$$\sup_{t\in[s,T-\delta]}\ \sup_{h\in[0,\delta]} |\langle g, \bar{\mu}^r(t+h)\rangle - \langle g, \bar{\mu}^r(t)\rangle| \le \tilde{\epsilon}. \tag{137}$$

In order to prove Theorem 5.2, we need Lemmas 5.3 and 5.4 below. The first lemma, Lemma 5.3, implies that if the fluid-scaled workload is bounded below, then so is the fluid-scaled queue length. This lemma (and its proof) is much like Gromoll et al. [9, Lemma 5.4]. The second lemma, Lemma 5.4, implies that in the presence of (135) and (136) there is an upper bound for the amount of mass that $\bar{\mu}^r(t)$ can have concentrated near the origin. This lemma (and its proof) is much like Gromoll et al. [9, Lemma 5.5]. These two lemmas are used below to prove Theorem 5.2. The respective proofs of Lemmas 5.3 and 5.4 are much like those of [9, Lemmas 5.4 and 5.5], except that time zero is replaced by a positive time s. Readers interested in this analysis are referred to [9].

Lemma 5.3. *Let $T > (4\alpha)^{-1}$ and $0 < \epsilon, \eta < 1$. Let T_ϵ, l, M_0, M_T, γ, K, Γ, κ, and r_0 be the constants, and $\{B^r\}_{r>0}$ be the events, given by Lemma 5.2. Suppose that $0 \le s < T - l$, $0 < r$, and C_s^r is an event such that $C_s^r \subset B^r$ and (136) holds on C_s^r. Then, on C_s^r,*

$$\inf_{t\in[s,T]} \langle 1, \bar{\mu}^r(t)\rangle \ge \frac{1}{\Gamma}. \tag{138}$$

Lemma 5.4. *Let $T > (4\alpha)^{-1}$ and $0 < \epsilon$, $\eta < 1$. Let T_ϵ, l, M_0, M_T, γ, K, Γ, κ, and r_0 be the constants, and $\{B^r\}_{r>0}$ be the events, given by Lemma 5.2. Suppose that $0 \le s < T - l$, $0 < r$, and C_s^r is an event such that $C_s^r \subset B^r$ and both (135) and (136) hold on C_s^r. Then, on C_s^r,*

$$\sup_{t\in[s,T]} \langle 1_{[0,\kappa]}, \bar{\mu}^r(t)\rangle \le \epsilon. \tag{139}$$

Proof of Theorem 5.2. First, note that $[s, T-\delta]$ is nonempty by (134) and since $s < T - l$. Fix $t \in [s, T-\delta]$ and $h \in [0, \delta]$. Observe that, by (138), on C_s^r,

$$\bar{S}^r_{t,t+h} \le \frac{h}{\inf_{u\in[s,T]}\langle 1, \bar{\mu}^r(u)\rangle} \le h\Gamma. \tag{140}$$

A first-order Taylor expansion of g gives the following estimate on C_s^r for all $y \in (\bar{S}^r_{t,t+h}, \infty)$:

$$|g(y - \bar{S}^r_{t,t+h}) - g(y)| = |-\bar{S}^r_{t,t+h} g'(w_y)| \le h\Gamma \|g'\|_\infty, \tag{141}$$

for some $w_y \in [y - \bar{S}^r_{t,t+h}, y]$, where the inequality follows by (140). Subtracting $\langle g, \bar{\mu}^r(t)\rangle$ from both sides of equation (115) and using the fact that $(1_{(0,\infty)}g)(\cdot - \bar{S}^r_{t,t+h}) = 1_{(\bar{S}^r_{t,t+h},\infty)}(\cdot) g(\cdot - \bar{S}^r_{t,t+h})$ yields that, on C_s^r,

$$\begin{aligned}
|\langle g, \bar{\mu}^r(t+h)\rangle - \langle g, \bar{\mu}^r(t)\rangle| &= \Bigg| \langle 1_{(\bar{S}^r_{t,t+h},\infty)}(\cdot)(g(\cdot - \bar{S}^r_{t,t+h}) - g(\cdot)), \bar{\mu}^r(t)\rangle - \langle 1_{[0,\bar{S}^r_{t,t+h}]} g, \bar{\mu}^r(t)\rangle \\
&\qquad + \frac{1}{r}\sum_{i=r\bar{E}^r(t)+1}^{r\bar{E}^r(t+h)} (1_{(0,\infty)}g)(v_i^r - \bar{S}^r_{U_i^r/r,t+h}) \Bigg| \\
&\le \langle |1_{(\bar{S}^r_{t,t+h},\infty)}(\cdot)(g(\cdot - \bar{S}^r_{t,t+h}) - g(\cdot))|, \bar{\mu}^r(t)\rangle + \|g\|_\infty \langle 1_{[0,h\Gamma]}, \bar{\mu}^r(t)\rangle \\
&\qquad + \|g\|_\infty (\bar{E}^r(t+h) - \bar{E}^r(t)) \\
&\le h\Gamma\|g'\|_\infty \langle 1, \bar{\mu}^r(t)\rangle + \|g\|_\infty \langle 1_{[0,h\Gamma]}, \bar{\mu}^r(t)\rangle + \|g\|_\infty (\bar{E}^r(t+h) - \bar{E}^r(t)),
\end{aligned}$$

where the first inequality is by (140) and the second is by (141). Now, taking the supremum over $h \in [0, \delta]$ and $t \in [s, T-\delta]$, we see that, on C_s^r,

$$\begin{aligned}
&\sup_{t\in[s,T-\delta]} \sup_{h\in[0,\delta]} |\langle g, \bar{\mu}^r(t+h)\rangle - \langle g, \bar{\mu}^r(t)\rangle| \\
&\quad \le \sup_{t\in[s,T-\delta]} (\delta\Gamma\|g'\|_\infty \langle 1, \bar{\mu}^r(t)\rangle + \|g\|_\infty \langle 1_{[0,\delta\Gamma]}, \bar{\mu}^r(t)\rangle + \|g\|_\infty (\bar{E}^r(t+\delta) - \bar{E}^r(t))) \\
&\quad \le \sup_{t\in[s,T-\delta]} \left(\frac{\tilde{\epsilon}}{4M_T} \langle 1, \bar{\mu}^r(t)\rangle + \|g\|_\infty \langle 1_{[0,\kappa]}, \bar{\mu}^r(t)\rangle + \|g\|_\infty (\bar{E}^r(t+l) - \bar{E}^r(t)) \right) \\
&\quad \le \frac{\tilde{\epsilon}}{4M_T} M_T + \|g\|_\infty \epsilon + \|g\|_\infty \frac{\epsilon}{4} \\
&\quad \le \frac{\tilde{\epsilon}}{4} + \frac{\tilde{\epsilon}}{2} + \frac{\tilde{\epsilon}}{8} < \tilde{\epsilon},
\end{aligned}$$

where the second inequality is by (134) and the third is by (132), (139), and (118). □

The next objective is to choose s and C_s^r for use in Theorem 5.2. We will choose $s = 0$ in combination with sets C_0^r where the initial workload is sufficiently large, and we will choose $s = T_\epsilon$ in combination with sets $C_{T_\epsilon}^r$ where the initial workload is small. For this, given $\gamma > 0$, for each $r > 0$, let

$$D_\gamma^r = \left\{ \langle \chi, \bar{\mu}^r(0)\rangle \le \frac{\gamma}{2} \right\},$$

and let $\breve{D}_\gamma^r$ be the complement of D_γ^r.

Lemma 5.5. *Let $T > (4\alpha)^{-1}$ and $0 < \epsilon, \eta < 1$. Let T_ϵ, l, M_0, M_T, γ, K, Γ, κ, and r_0 be the constants, and $\{B^r\}_{r>0}$ be the events, given by Lemma 5.2. On $B^r \cap D^r_\gamma$,*

$$\inf_{t\in[T_\epsilon, T]} \langle \chi, \bar{\mu}^r(t)\rangle \geq \frac{\gamma}{4}. \tag{142}$$

Also, on $B^r \cap \check{D}^r_\gamma$,

$$\inf_{t\in[0, T]} \langle \chi, \bar{\mu}^r(t)\rangle \geq \frac{\gamma}{4}. \tag{143}$$

Proof. By using (133), (125), and the definition of D^r_γ,

$$\langle \chi, \bar{\mu}^r(t)\rangle \geq (\rho-1)T_\epsilon - \frac{\gamma}{4} > \frac{\gamma}{2} - \frac{\gamma}{4} = \frac{\gamma}{4}, \quad t\in[T_\epsilon, T], \text{ on } B^r\cap D^r_\gamma,$$
$$\langle \chi, \bar{\mu}^r(t)\rangle \geq \frac{\gamma}{2} - \frac{\gamma}{4} = \frac{\gamma}{4}, \quad t\in[0, T], \text{ on } B^r\cap \check{D}^r_\gamma. \quad \square$$

Lemma 5.5 implies that (136) holds on $B^r\cap \check{D}^r_\gamma$ for $s=0$ and on $B^r\cap D^r_\gamma$ for $s=T_\epsilon$. On $B^r\cap \check{D}^r_\gamma$, (135) holds for $s=0$; this is an immediate consequence of (124). On the other hand, as a consequence of the next lemma, (135) holds on $B^r\cap D^r_\gamma$ for $s=T_\epsilon$.

Lemma 5.6. *Let $T > (4\alpha)^{-1}$ and $0 < \epsilon$, $\eta < 1$. Let T_ϵ, l, M_0, M_T, γ, K, Γ, κ, and r_0 be the constants, and $\{B^r\}_{r>0}$ be the events, given by Lemma 5.2. Then, on $B^r \cap D^r_\gamma$,*

$$\langle 1, \bar{\mu}^r(0)\rangle \leq \frac{\epsilon}{4} \tag{144}$$

and

$$\sup_{t\in[0, T_\epsilon]} \langle 1, \bar{\mu}^r(t)\rangle \leq \frac{\epsilon}{2}. \tag{145}$$

Proof. On $B^r\cap D^r_\gamma$, by (124), the definition of D^r_γ, and (125),

$$\begin{aligned}\langle 1, \bar{\mu}^r(0)\rangle &= \langle 1_{[0,\kappa]}, \bar{\mu}^r(0)\rangle + \langle 1_{(\kappa,\infty)}, \bar{\mu}^r(0)\rangle\\ &\leq \frac{\epsilon}{8} + \frac{1}{\kappa}\langle \chi, \bar{\mu}^r(0)\rangle\\ &\leq \frac{\epsilon}{8} + \frac{\gamma}{2\kappa}\\ &\leq \frac{\epsilon}{4}.\end{aligned}$$

By (131), (144), (119), and (116), on $B^r\cap D^r_\gamma$,

$$\sup_{t\in[0, T_\epsilon]} \langle 1, \bar{\mu}^r(t)\rangle \leq \frac{\epsilon}{4} + 2\alpha T_\epsilon = \frac{\epsilon}{2}. \quad \square$$

We are now ready to verify conditions sufficient to imply the tightness of $\{\langle g, \bar{\mu}^r(\cdot)\rangle\}_{r>0}$, for each $g\in \mathbf{C}^1_b(\mathbb{R}_+)$.

Theorem 5.3. *Let $g\in \mathbf{C}^1_b(\mathbb{R}_+)$, $T>0$, and $0<\tilde{\epsilon}, \eta<1$. Then, there exist M, $\delta>0$, and $r_0>0$, such that $r>r_0$ implies*

$$\mathbf{P}\left(\sup_{t\in[0,T]} |\langle g, \bar{\mu}^r(t)\rangle| \leq M\right) \geq 1-\eta, \tag{146}$$

$$\mathbf{P}\left(\sup_{t\in[0,T-\delta]}\sup_{h\in[0,\delta]} |\langle g, \bar{\mu}^r(t+h)\rangle - \langle g, \bar{\mu}^r(t)\rangle| \leq \tilde{\epsilon}\right) \geq 1-\eta. \tag{147}$$

Proof. Define $\epsilon = \tilde{\epsilon}/2(\|g\|_\infty \vee 1)$. Without loss of generality, assume that $T > (4\alpha)^{-1}$. Then, by Lemma 5.2, there exist constants T_ϵ, l, M_0, M_T, γ, K, Γ, κ, and r_0, and events $\{B^r\}_{r>0}$, such that $r>r_0$ implies $\mathbf{P}(B^r)\geq 1-\eta$, and on B^r for all $r>0$, (116)–(132) hold. Let δ be given by (134) and define

$$M = (\|g\|_\infty \vee 1)M_T.$$

To prove (146), observe that on B^r, by (132),

$$\begin{aligned}\sup_{t\in[0,T]}|\langle g,\bar{\mu}^r(t)\rangle| &\le \|g\|_\infty \sup_{t\in[0,T]}\langle 1,\bar{\mu}^r(t)\rangle\\ &\le \|g\|_\infty M_T\\ &\le M.\end{aligned}$$

Next, we prove (147). Let $s=0$ and $C_0^r=B^r\cap\breve{D}_\gamma^r$. Then, (124) implies (135) and (143) implies (136). So, by Theorem 5.2, (137) holds on $B^r\cap\breve{D}_\gamma^r$ for all $r>0$ with $s=0$. For $s=T_\epsilon$ and $C_{T_\epsilon}^r=B^r\cap D_\gamma^r$, (145) implies (135) and (142) implies (136). So, by Theorem 5.2, (137) holds on $B^r\cap D_\gamma^r$ for all $r>0$ with $s=T_\epsilon$. Also on $B^r\cap D_\gamma^r$,

$$\begin{aligned}\sup_{t\in[0,T_\epsilon]}\sup_{h\in[0,\delta]}|\langle g,\bar{\mu}^r(t+h)\rangle-\langle g,\bar{\mu}^r(t)\rangle| &\le 2\sup_{t\in[0,T_\epsilon+\delta]}|\langle g,\bar{\mu}^r(t)\rangle|\\ &\le 2\|g\|_\infty\sup_{t\in[0,2T_\epsilon]}\langle 1,\bar{\mu}^r(t)\rangle\\ &\le 2\|g\|_\infty\left(\frac{\epsilon}{4}+4\alpha T_\epsilon\right)\\ &= 2\|g\|_\infty\frac{3\epsilon}{4}\le\tilde{\epsilon}.\end{aligned}\tag{148}$$

Here, the second line uses $\delta\le l$ and (117), and the third line uses (131), (144), and (119). The fourth line uses (116). Thus, the desired estimate holds on B^r. Since, for $r>r_0$, $\mathbf{P}(B^r)\ge 1-\eta$, (147) holds. □

Proof of Tightness for Theorem 5.1. Recall that it suffices to show conditions (T.1) and (T.2) hold. Let $T>0$, $0<\tilde{\epsilon}$, $\eta<1$. Define $N_T=\max\{M_T,\gamma/4+M_0+(\rho-1)T\}$. To show (T.1), define

$$C_{T,\eta}=\{\zeta\in\mathcal{M}_F\colon\langle 1,\zeta\rangle\vee\langle\chi,\zeta\rangle\le N_T\}.$$

Since $\langle\chi,\zeta\rangle\le N_T$ implies $\langle 1_{[K,\infty)},\zeta\rangle\le N_T/K$, we have

$$\sup_{\zeta\in C_{T,\eta}}\langle 1_{[K,\infty)},\zeta\rangle\longrightarrow 0,$$

as $K\to\infty$, which implies that $C_{T,\eta}\subset\mathcal{M}_F$ is relatively compact (cf. Kallenberg [11, Theorem 15.7.5]). By (132), (133), and (120),

$$\liminf_{r\to\infty}\mathbf{P}(\bar{\mu}^r(t)\in C_{T,\eta}\text{ for all }t\in[0,T])\ge 1-\eta.$$

Replace $C_{T,\eta}$ with its closure and (T.1) is proved. Finally, (T.2) follows directly from Theorem 5.3 by applying a standard tightness criterion for real-valued processes (cf. Ethier and Kurtz [5, Chapter 3, Corollary 7.4]). □

5.3. Proof of limit point properties. In this section, we complete the proof of Theorem 5.1 by showing that any limit point of the sequence $\{\bar{\mu}^r(\cdot)\}_{r>0}$ is a.s. a fluid model solution for the strictly supercritical data (α,ν). In particular, we show that a.s., each sample path is continuous (cf. (S.1)), has no atoms at the origin for all time (cf. (S.2)), and is a solution of the strictly supercritical fluid model equations (cf. (S.3)). In order to prove (S.3), we first show that a.s. the sample paths have positive mass at all positive times (cf. (21)), which eliminates the need for $\varphi(\cdot)$ in (16) and simplifies the proof. The proofs of (S.1)–(S.3) are modifications of those in Gromoll et al. [9] for (C.1)–(C.4). Care is needed in adapting these proofs when the initial measure for the limit point is the zero measure.

For this, fix a limit point $\bar{\mu}^\star(\cdot)$ of the sequence $\{\bar{\mu}^r(\cdot)\}_{r>0}$. Since $\bar{\mu}^\star(\cdot)$ is a limit point of $\{\bar{\mu}^r(\cdot)\}_{r>0}$, there exists a subsequence $\{\bar{\mu}^{r'}(\cdot)\}_{r'>0}\subset\{\bar{\mu}^r(\cdot)\}_{r>0}$ such that $\bar{\mu}^{r'}(\cdot)\Rightarrow\bar{\mu}^\star(\cdot)$ as $r'\to\infty$. To ease notation for the remainder of the proof, we relabel r' as r, remembering that we have passed to a subsequence which converges in distribution to $\bar{\mu}^\star(\cdot)$.

5.3.1. Continuity of limit points: Verification of (S.1). From Theorem 5.3, and more specifically from (147), it follows that a.s., $\bar{\mu}^\star(\cdot)$ has continuous sample paths. To see this, choose a countable set $V\subset\mathbf{C}_b^1(\mathbb{R}_+)$ that separates elements of $\mathcal{M}_F$ (cf. Ethier and Kurtz [5, Chapter 3, Proposition 4.2.ff]). By (147), for each $g\in V\subset\mathbf{C}_b^1(\mathbb{R}_+)$, the real-valued process $\langle g,\bar{\mu}^\star(\cdot)\rangle$ a.s. has continuous sample paths. Since V is a countable separating class for $\mathcal{M}_F$, the result follows.

5.3.2. Positive total mass and no atom at the origin for all positive times: Verification of (21) and (S.2). Here, it is shown that for each $T > (4\alpha)^{-1}$, a.s.,

$$\langle 1_{\{0\}}, \bar{\mu}^{\star}(t)\rangle = 0, \quad \text{for all } t \in [0, T), \tag{149}$$

$$\bar{\mu}^{\star}(t) \neq \mathbf{0}, \quad \text{for all } t \in (0, T). \tag{150}$$

Note that by (52), $\langle 1_{\{0\}}, \bar{\mu}^{\star}(0)\rangle = 0$. Thus, in order to prove (149) and (150), it suffices to show that for each $0 < S < T$, a.s.,

$$\langle 1_{\{0\}}, \bar{\mu}^{\star}(t)\rangle = 0, \quad \text{for all } t \in (S, T), \tag{151}$$

$$\bar{\mu}^{\star}(t) \neq \mathbf{0}, \quad \text{for all } t \in (S, T). \tag{152}$$

The proof of (151) and (152) is a simple modification of the proof of Gromoll et al. [9, Equations (5.51)–(5.53)], which we now explain. For this, let $\mathbf{Q}^r$ and $\mathbf{Q}^{\star}$ be the probability laws induced on $D([0,\infty), \mathcal{M}_F)$ by $\bar{\mu}^r(\cdot)$ and $\bar{\mu}^{\star}(\cdot)$, respectively. Fix $0 < S < T$ and define the sets A' and A''' as in [9], except replace the time interval $[0, T)$ by (S, T):

$$A' = \left\{\zeta(\cdot) \in D([0,\infty), \mathcal{M}_F)\colon \sup_{t\in(S,T)} \langle 1_{\{0\}}, \zeta(t)\rangle > 0\right\},$$

$$A''' = \left\{\zeta(\cdot) \in D([0,\infty), \mathcal{M}_F)\colon \inf_{t\in(S,T)} \langle 1, \zeta(t)\rangle = 0\right\}.$$

Let $A = A' \cup A'''$. Clearly, any sample path of $\bar{\mu}^{\star}(\cdot)$ which is not an element of A satisfies (151) and (152). So, it suffices to show that A is contained in a $\mathbf{Q}^{\star}$-null set.

In order to show that A is contained in a $\mathbf{Q}^{\star}$-null set, the Borel-Cantelli argument in Gromoll et al. [9] is adapted to the present situation. To do this, for each $n \in \mathbb{N}$, choose a pair $0 < \epsilon_n, \eta_n < 1$ as in [9], except we also require that $\epsilon_n < 8\alpha S$; that is, we require that $\sum_{n=1}^{\infty} \eta_n < \infty$, $\epsilon_n < 8\alpha S$ for each $n \in \mathbb{N}$, and $(\epsilon_n, \eta_n) \to (0, 0)$ as $n \to \infty$. Then, by Lemma 5.2, for each $n \in \mathbb{N}$ there exist strictly positive constants T_{ϵ_n}, l_n, $M_{0,n}$, $M_{T,n}$, κ_n, γ_n, K_n, Γ_n, and $r_{0,n}$, and events $\{B_n^r\}_{r>0}$ such that $r > r_{0,n}$ implies $\mathbf{P}(B_n^r) \geq 1 - \eta_n$, and (116)–(128) hold on B_n^r with the above constants in place of the analogous ones appearing there. Note that the additional requirement on ϵ_n implies that $T_{\epsilon_n} < S$. Using (142), (145), and Lemmas 5.3 and 5.4 for $s = T_{\epsilon_n}$ and $C^r_{T_{\epsilon_n}} = B_n^r \cap D^r_{\gamma_n}$, using (124), (143), and Lemmas 5.3 and 5.4 for $s = 0$ and $C_0^r = B_n^r \cap \check{D}^r_{\gamma_n}$, and using the fact that $T_{\epsilon_n} < S$, it follows that, for each $r > 0$, $n \in \mathbb{N}$, and $0 < c_n < \kappa_n$, on the event B_n^r,

$$\inf_{t\in(S,T)} \langle 1, \bar{\mu}^r(t)\rangle \geq \inf_{t\in[T_{\epsilon_n},T]} \langle 1, \bar{\mu}^r(t)\rangle \geq \frac{1}{\Gamma_n}, \tag{153}$$

$$\sup_{t\in(S,T)} \langle 1_{[0,c_n)}, \bar{\mu}^r(t)\rangle \leq \sup_{t\in[T_{\epsilon_n},T]} \langle 1_{[0,c_n)}, \bar{\mu}^r(t)\rangle \leq \epsilon_n. \tag{154}$$

Fix $\{c_n\}_{n\in\mathbb{N}}$ such that $c_n \to 0$ as $n \to \infty$. For each $n \in \mathbb{N}$, let

$$A_n' = \left\{\zeta(\cdot) \in D([0,\infty), \mathcal{M}_F)\colon \sup_{t\in(S,T)} \langle 1_{[0,c_n)}, \zeta(t)\rangle > \epsilon_n\right\},$$

$$A_n''' = \left\{\zeta(\cdot) \in D([0,\infty), \mathcal{M}_F)\colon \inf_{t\in(S,T)} \langle 1, \zeta(t)\rangle < \frac{1}{\Gamma_n}\right\}.$$

Then, use a Borel-Cantelli argument as in [9] to complete the proof.

5.3.3. The derivation of the strictly supercritical fluid model equations: Verification of (S.3). The derivation of (16) for $\bar{\mu}^{\star}(\cdot)$ proceeds in a manner nearly identical to the derivation of Gromoll et al. [9, Equation (3.3)] (cf. [9, p. 849, Proof of Property (3)]). The essential difference is that time zero is replaced by a positive time. In particular, by (150), it suffices to show that, a.s. for all $g \in \mathcal{C}$ and $0 < s \leq t < \infty$,

$$\langle g, \bar{\mu}^{\star}(t)\rangle = \langle g, \bar{\mu}^{\star}(s)\rangle + \int_s^t \frac{\langle g', \bar{\mu}^{\star}(u)\rangle}{\langle 1, \bar{\mu}^{\star}(u)\rangle}\, du + \alpha(t-s)\langle g, \nu\rangle. \tag{155}$$

For this, fix $T > (4\alpha)^{-1}$ and $0 < S < T < \infty$, and for each n choose a pair (ϵ_n, η_n) as in §5.3.2. Then, to obtain (155) for each $S \leq s \leq t \leq T$, proceed as in the derivation of [9, Equation (3.3)], except replacing time zero with time s and using (153) above in place of [9, Equation (5.55)].

5.4. Proof of convergence to fluid model solutions.

PROOF OF THEOREM 3.7. By Theorem 5.1, the sequence $\{\bar{\mu}^r(\cdot)\}_{r>0}$ of measure-valued processes is tight, and any limit point $\bar{\mu}^\star(\cdot)$ has sample paths which a.s. are fluid model solutions for the strictly supercritical data (α, ν). By (50), $\bar{\mu}^\star(0)$ is equal in distribution to Θ, so it remains to show that $\bar{\mu}^\star(\cdot)$ is unique in law. Theorem 3.1 asserts that fluid model solutions for the data (α, ν) are unique given an initial value $\xi \in \mathcal{M}_F^c$. More precisely, given a $\xi \in \mathcal{M}_F^c$, the unique fluid model solution $\bar{\mu}(\cdot)$ for critical data (α, ν) with initial value $\bar{\mu}(0) = \xi$ is given by $\Xi(\xi) = \bar{\mu}_\xi(\cdot)$. Since $\bar{\mu}^\star(0)$ is equal in distribution to Θ, we have by (52) that $\bar{\mu}^\star(0) \in \mathcal{M}_F^c$ a.s. Thus, we see that a.s.,

$$\bar{\mu}^\star(\cdot) = \Xi(\bar{\mu}^\star(0)). \tag{156}$$

By Theorem 3.3, the mapping Ξ is continuous. So, the law of $\bar{\mu}^\star(\cdot)$ is uniquely determined by the law of the random measure Θ, which completes the proof. □

Acknowledgments. The research of Amber Puha was supported in part by a National Science Foundation (NSF) Mathematical Sciences Postdoctoral Research Fellowship. The research of Ruth Williams was supported in part by NSF Grants DMS 0071408 and 0305272, a John Simon Guggenheim Fellowship, and a gift from the David and Holly Mendel fund.

References

[1] Billingsley, P. 1968. *Convergence of Probability Measures*. John Wiley & Sons, Inc., New York.
[2] Chen, H., O. Kella, G. Weiss. 1997. Fluid approximations for a processor-sharing queue. *Queueing Systems: Theory Appl.* **27** 99–125.
[3] Doytchinov, B., J. Lehoczky, S. Shreve. 2001. Real-time queues in heavy traffic with earliest-deadline-first-queue discipline. *Ann. Appl. Probab.* **11** 332–378.
[4] Durrett, R. T. 1996. *Probability: Theory and Examples*, 2nd ed. Duxbury Press, Belmont, CA.
[5] Ethier, S. N., T. G. Kurtz. 1986. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, Inc., New York.
[6] Feller, W. 1971. *An Introduction to Probability Theory and its Applications*, Vol. 2, 2nd ed. John Wiley & Sons, Inc., New York.
[7] Grishechkin, S. 1994. GI/G/1 processor sharing queue in heavy traffic. *Adv. Appl. Probab.* **26** 539–555.
[8] Gromoll, H. C. 2004. Diffusion approximation for a processor sharing queue in heavy traffic. *Ann. Appl. Probab.* **14** 555–611.
[9] Gromoll, H. C., A. L. Puha, R. J. Williams. 2002. The fluid limit of a heavily loaded processor sharing queue. *Ann. Appl. Probab.* **12** 797–859.
[10] Jean-Marie, A., P. Robert. 1994. On the transient behavior of the processor sharing queue. *Queueing Systems: Theory Appl.* **17** 129–136.
[11] Kallenberg, O. 1983. *Random Measures*. Academic Press, New York.
[12] Linz, P. 1985. *Analytical and Numerical Methods for Volterra Equations*, *SIAM Studies in Applied Mathematics*, Vol. 7. SIAM, Philadelphia, PA.
[13] Mandjes, M., B. Zwart. Large deviations and importance sampling for sojourn times in processor sharing queues. *Queueing Systems: Theory Appl.* Forthcoming.
[14] Prohorov, Yu V. 1956. Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.* **1** 157–214.
[15] Puha, A. L., R. J. Williams. 2004. Invariant states and rates of convergence for the fluid limit of a heavily loaded processor sharing queue. *Ann. Appl. Probab.* **14** 517–554.
[16] Resnick, S. I. 1992. *Adventures in Stochastic Processes*. Birkhauser, Boston, MA.
[17] Smith, J. M. 1989. *Evolutionary Genetics*. Oxford University Press, New York.