Bernoulli **25**(4B), 2019, 3734–3761
https://doi.org/10.3150/19-BEJ1109

Estimating the input of a Lévy-driven queue by Poisson sampling of the workload process

LIRON RAVNER[1,2], ONNO BOXMA[2] and MICHEL MANDJES[1]

[1]*University of Amsterdam, Korteweg-de Vries Institute for Mathematics, Science Park, Amsterdam, 1098XG, Netherlands. E-mail: l.ravner@uva.nl*
[2]*Eindhoven University of Technology, Department of Mathematics and Computer Science, 5600MB, Eindhoven, Netherlands.*

This paper aims at semi-parametrically estimating the input process to a Lévy-driven queue by sampling the workload process at Poisson times. We construct a method-of-moments based estimator for the Lévy process' characteristic exponent. This method exploits the known distribution of the workload sampled at an exponential time, thus taking into account the dependence between subsequent samples. Verifiable conditions for consistency and asymptotic normality are provided, along with explicit expressions for the asymptotic variance. The method requires an intermediate estimation step of estimating a constant (related to both the input distribution and the sampling rate); this constant also features in the asymptotic analysis. For subordinator Lévy input, a partial MLE is constructed for the intermediate step and we show that it satisfies the consistency and asymptotic normality conditions. For general spectrally-positive Lévy input a biased estimator is proposed that only uses workload observations above some threshold; the bias can be made arbitrarily small by appropriately choosing the threshold.

Keywords: Lévy-driven queue; nonparametric estimation; Poisson probing; transient queueing; queue input estimation

1. Introduction

To optimally design and control queueing systems, it is of great importance to have reliable estimates of the model primitives such as the arrival rate and the service-time distribution. In many situations, however, one cannot observe the users' interarrival times and service times; instead, one only has periodic observations of the workload process. This leaves us with the challenging task of inferring the model primitives from such workload observations. A complicating factor is that subsequent observations are dependent. As a consequence, standard estimation techniques (such as maximum-likelihood estimation) are not applicable in almost all queuing systems.

This work focuses on the problem described above in the setting of a single-server queue with work arriving according to a *spectrally-positive Lévy* inputs process, i.e., a Lévy process $X(\cdot)$ with only positive jumps. The resulting model includes the M/G/1 queue as a special case; then the input process is a Compound Poisson (CP) process. The set of spectrally-positive Lévy input processes, however, is substantially broader than CP: it covers a broad range of other relevant processes, such as Brownian motion, the Gamma process and the Inverse Gaussian process. The Lévy input process is uniquely defined by its *characteristic exponent* function $\varphi(\alpha) := \log \mathrm{E} e^{-\alpha X(1)}$, with $\alpha \geq 0$ [27]. We assume that $\varphi(\cdot)$ is unknown and it is therefore the objective to estimate this function. For a detailed account of Lévy-driven queues we refer to [11].

1350-7265 © 2019 ISI/BS

Approach. In our approach, we exploit the fact that, given some initial workload, there is an explicit expression for the Laplace–Stieltjes Transform (LST) of the workload after an exponentially distributed time; see, for example, [26]. When the workload is sampled at random times according to an independent Poisson process (with some specified rate ξ), this LST enables us to deal with the dependence between the subsequent workload observations. In particular, method of moments estimation is carried out by equating the empirical and conditional expected LST, thus obtaining so called *Z-estimators* [35]. Our setting can be viewed as non-parametric in the sense that we do not make specific distributional assumptions about the input process except that it is a spectrally-positive Lévy process. However, we use parametric estimation techniques for pointwise estimation of the exponent function.

There is the complication that, due to the specific form of the above-mentioned LST, it is necessary to first estimate $\varphi^{-1}(\xi)$, that is, the inverse of φ at the sampling rate ξ. This intermediate step is also crucial for determining the asymptotic performance of the estimator for $\varphi(\alpha)$. In particular, we show that consistency and asymptotic normality in the intermediate step (to estimate $\varphi^{-1}(\xi)$, that is) carry over to consistency and asymptotic normality in the main estimation step (to estimate $\varphi(\alpha)$). For the case of subordinator input (i.e., a Lévy process whose paths are non-decreasing almost surely) we construct a maximum-likelihood procedure for estimating $\varphi^{-1}(\xi)$, which we show to be consistent and asymptotically normal. This result therefore covers the special cases of the driving Lévy process being CP, a Gamma process and an Inverse Gaussian process. For general spectrally-positive Lévy input, that is, also including a Brownian motion component, we provide an estimation method for $\varphi^{-1}(\xi)$ that uses workload observations only above some threshold; this procedure is inherently biased, but the bias can be controlled by choosing the threshold parameter appropriately.

Later in this Introduction, we describe in detail the differences with existing estimation approaches. A major difference is that previous papers tend to aim at using workload observations to estimate quantities pertaining to the steady-state workload distribution, from which the model primitives are then inferred, whereas we explicitly use the knowledge of the dependence between two subsequent workload observations.

Main contributions. We now summarize the main contributions of our work.

- Thus far, the literature has mostly focused on the case of Compound Poisson input, whereas we address the larger class of spectrally-positive Lévy processes.
- In our approach we do not rely on estimating quantities pertaining to the steady-state workload distribution, but rather explicitly use the system's correlation structure.
- We provide verifiable conditions for consistency and asymptotic normality of our estimator (including an expression for the asymptotic variance). These conditions are verified for the case of subordinator input, thus implying the asymptotic properties for the special case of CP. When $\varphi(\alpha)$ is estimated for multiple values of α simultaneously, we provide the corresponding asymptotic covariances.
- The idea to use a Z-estimator in this Lévy-driven setting is novel. As mentioned, our approach requires knowledge of $\varphi^{-1}(\xi)$, for which we present an estimation procedure.
- We present simulation examples that explore the accuracy of the procedure. These show that the accuracy is generally good, but degrades for large α (because the asymptotic variance of the error increases with α). We also observe that the accuracy is best for low ξ, which is to be expected: when ξ is high, the workload process is very frequently sampled, such that new

observations offer hardly any new information. As mentioned, for non-subordinator input the estimator of $\varphi^{-1}(\xi)$ is biased, but our experiments show that the accuracy is high, even for a relatively low threshold.

Background and related literature. Observing a system at Poisson times is commonly known as *Poisson probing*. In communication networks it has been used to estimate (steady-state) packet delays in a network with unknown traffic input [2,34]. In a practical implementation of Poisson probing, very small jobs are sent through the system and their delays are recorded. This method relies on the *Poisson Arrivals See Time Averages* (PASTA) property: the stationary distribution coincides with the distribution observed at Poisson instants; theoretical justification can be found in, for example, [14,36].

In the M/G/1 case, the queue is fed by a Poisson arrival process with rate λ, whereas the service-times are i.i.d. random variables $G_i \sim G$ with LST G^*. In [8], a non-parametric approach is proposed for estimating λ and the LST of G using observations of the lengths of the busy and idle periods. This procedure overcomes the issue of dependent observations as busy periods are independent, but it may require a very long observation period to get a large sample of busy periods, especially in heavily loaded systems. The estimator exploits the fixed-point equation $B^*(\alpha) = G^*(\alpha + \lambda(1 - B^*(\alpha)))$, where B^* is the LST of the busy period, and uses an estimator $\hat{G}_n(\alpha)$ based on the empirical LST of the busy periods (with n observations). The estimator is consistent and asymptotically normal: $\sqrt{n}(\hat{G}_n(\cdot) - G^*(\cdot))$ converges in distribution to a Gaussian process. Reference [20] provides a nonparametric procedure for estimating the service-time distribution, using an equidistant sample of the workload. It consists of two steps: (1) estimating the residual service-time distribution by inverting the convolution workload formula, $F_V = (1-\rho)\sum_{k=0}^{\infty} \rho^k G_e^{\star k}$, where F_V is the stationary cdf of the workload and $G_e^{\star k}$ is the k-fold convolution of the residual service-time distribution, and (2) then estimating the service-time distribution itself. As the first step relies on the stationary workload distribution, the procedure does not use any knowledge on the dependence between observations. Nevertheless, the estimation of the residual service-time distribution is shown to be consistent and asymptotically normal under some conditions on the system load. For the second step only heuristic methods are suggested. The numerical experiments in [20] show that the procedure performs well for moderately loaded systems but often performs poorly for systems with low or high load.

A general framework for generalized method-of-moments (GMM) estimators for Markov processes sampled according to a Poisson process, that is possibly state and time dependent, is provided in [12]. As it turns out, the fact that in our setup the quantity $\varphi^{-1}(\xi)$ has to be estimated entails that we cannot use this approach. In Section 7, we further discuss the relation of our work with the GMM-based procedure of [12] in the context of our model.

Non-parametric estimation of the stationary waiting time distribution for a GI/G/1 queue was studied in [30]. If the number of arrivals during a service time can be observed, then estimation of the service-time distribution reduces to the problem of decompounding Poisson sums, as studied in [10]. As is often the case, when moving from a single-server queue to its infinite-server counterpart, the analysis greatly simplifies. The object of interest in this line of research is typically the service-time distribution, and we refer the reader to [8,9,15,16,33]. A related approach appeared in [29] where an estimator for the arrival rate is derived by relying on counting the number of arrivals during a busy period. In [32] diffusion approximations are used.

Estimation of the model primitives from queue observations also has important managerial applications, for example, in the contexts of demand estimation [4] and dynamic pricing [1]. In [3], the traffic intensity of a non-homogeneous-in-time queue was estimated using a (non-Poisson) probing scheme that depends on the service-time of the probes. This list of papers on statistical inference of queueing systems is by no means exhaustive; see [5] for an extensive bibliography of this strand of research.

In the present paper, we estimate (the characteristic exponent of) a Lévy process by observing a Lévy-driven queue. Various papers consider the counterpart in which the Lévy process itself is observed. We refer to, for example, [7,28]; [17] studies the case of a Lévy process without small jumps.

Paper organization. The remainder of the paper is organized as follows. Section 2 introduces the queueing model and sampling scheme. In Section 3, the estimator for $\varphi(\alpha)$ is defined and conditions for consistency and asymptotic normality are provided. The following two sections deal with the intermediate estimation step of $\varphi^{-1}(\xi)$. In Section 4 an MLE is constructed for the subordinator case and it is shown to satisfy the consistency and asymptotic normality conditions. Section 5 presents an approach for the intermediate step that relies only on workload observations above some threshold. Further, a bound is derived for the bias and explicitly computed in the limit when the sample size grows. In Section 6, we numerically analyze the performance of the estimators using simulated data. Section 7 contains conclusions and suggestions for further research. Finally, proofs and further technical details are provided in the appendices [31].

2. Model and preliminaries

We study a queue with a spectrally-positive Lévy input process $J(t)$. The distribution of the process can be represented by the exponent function,

$$\log \mathrm{E}e^{-\alpha J(1)} = -c\alpha + \frac{1}{2}\sigma^2\alpha^2 - \int_{(0,\infty)} \left(1 - e^{-\alpha x}\right)\nu(\mathrm{d}x), \tag{1}$$

where c and σ are non-negative constants and ν is the Lévy jump measure. The output of the system is a unit-rate negative linear drift. This means that the net input process is $X(t) = J(t) - t$, which is also a spectrally-positive Lévy process. For example, if $J(t)$ is CP with rate λ and jump size LST G^*, then the exponent function of the net input is $\varphi(\alpha) = \lambda(G^*(\alpha) - 1) + \alpha$.

Our objective is to estimate the (unknown) exponent function of the net input process, $\varphi(\alpha) := \log \mathrm{E}e^{-\alpha X(1)}$, with $\alpha \geq 0$. Let $(V(t))_{t\in\mathbb{R}}$ denote the workload process, which can be represented as the net input process reflected at zero: $V(t) = X(t) + \max\{V(0), L(t)\}$, where $L(t) := -\inf_{0\leq s\leq t} X(s)$. If $\mathrm{E}X(1) < 0$, then the stationary distribution of $V := V(\infty)$ is given by the *generalized Pollaczek–Khintchine* (GPK) formula [11], p. 27:

$$\mathrm{E}e^{-\alpha V} = \frac{\alpha\varphi'(0)}{\varphi(\alpha)}. \tag{2}$$

Furthermore, if $\sigma = 0$ then the input process $J(t)$ is a subordinator (i.e., an almost surely non-decreasing Lévy process), and the stationary probability that the workload is 0 equals

$$\mathrm{P}(V = 0) = \varphi'(0) = -\mathrm{E}X(1). \tag{3}$$

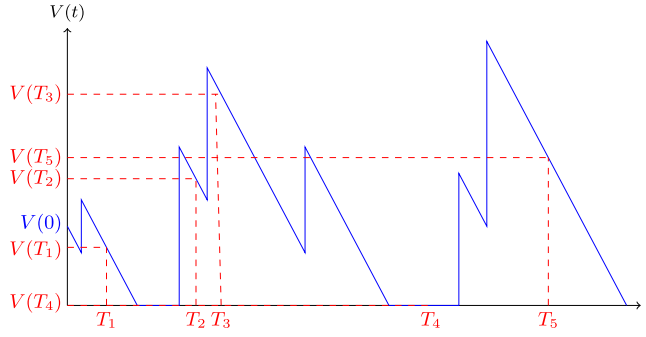

Figure 1. The workload $V(t)$ of an M/G/1 queue is sampled according to the probing process at times $T_1, T_2, \ldots$. The jump sizes and interarrival times are not observed. Inference is made based on the sample $(V(T_1), V(T_2), \ldots)$.

This is the is case if, for example, the input is CP (as in the M/G/1 model). Importantly, however, it also holds for spectrally-positive Lévy input without a Brownian component, that is, a Lévy jump process with non-negative jumps (covering e.g., the Gamma process and the Inverse Gaussian process).

All asymptotic results presented in this paper require that $\mathrm{E}X(1) < 0$, but in principle the underlying estimation procedure can still be performed when this assumption is not valid. In relation to the case of an unstable queue (i.e., if $\mathrm{E}X(1) \geq 0$), it is noted that, while the estimators are anticipated to perform reasonably well, other estimation procedures are likely to be more appropriate due to the fact that (after an initial transient period) one effectively observes the (non-reflected) net input process $(X(t))_{t\in\mathbb{R}}$.

From now on, we assume that the linear drift component c of the input process $(J(t))_{t\in\mathbb{R}}$ is known. Without loss of generality, we normalize this rate to $c = 0$, so that the linear drift of the net input process $(X(t))_{t\in\mathbb{R}}$ is -1. If both of the net input process and net output rate are unknown then there are evident problems of identification, and the estimation problem is not well defined; such cases are outside the scope of this paper.

Suppose that the workload is observed at n random times, sampled according to an independent Poisson process with rate ξ. Denote these times by $\{T_i\}_{i=1}^n$ such that $T_i - T_{i-1}$ are i.i.d. samples from an exponential distribution with parameter ξ (putting $T_0 := 0$). The sampling procedure is illustrated for the workload of an M/G/1 queue in Figure 1.

The distribution of the workload at T_i conditional on the workload at sample T_{i-1} is

$$\mathrm{E}\big[e^{-\alpha V(T_i)}|V(T_{i-1})\big] = \frac{\xi}{\xi - \varphi(\alpha)}\left(e^{-\alpha V(T_{i-1})} - \frac{\alpha}{\psi(\xi)}e^{-\psi(\xi)V(T_{i-1})}\right), \tag{4}$$

see, e.g. [26] and [11], Ch. IV; here ψ is the inverse of φ. For $\alpha = \psi(\xi)$, the LST of (4) can still be computed by applying L'Hôpital's rule:

$$\mathrm{E}\big[e^{-\psi(\xi)V(T_i)}|V(T_{i-1})\big] = \frac{\xi e^{-\psi(\xi)V(T_{i-1})}(V(T_{i-1}) + \frac{1}{\psi(\xi)})}{\varphi'(\psi(\xi))}. \tag{5}$$

Noting that $\varphi(0) = 0$, it can be verified that the conditional expected workload at observation time T_i is

$$\mathrm{E}\big(V(T_i)|V(T_{i-1})\big) = V(T_{i-1}) + \frac{e^{-\psi(\xi)V(T_{i-1})}}{\psi(\xi)} - \frac{\varphi'(0)}{\xi}, \tag{6}$$

and the conditional second moment is

$$\mathrm{E}\big[V(T_i)^2|V(T_{i-1})\big] = V(T_{i-1})^2 - 2\frac{\varphi'(0)}{\xi}\left(V(T_{i-1}) + \frac{e^{-\psi(\xi)V(T_{i-1})}}{\psi(\xi)} - \frac{\varphi'(0)}{\xi}\right) + \frac{\varphi''(0)}{\xi}. \tag{7}$$

The above holds for any Lévy-driven queue with spectrally-positive input. In the subordinator case, essentially due to the fact that the workload process spends time at zero with positive probability, by taking $\alpha \to \infty$ in (4), it turns out that

$$\mathrm{P}\big(V(T_i) = 0|V(T_{i-1}) = v\big) = \frac{\xi}{\psi(\xi)}e^{-\psi(\xi)v}. \tag{8}$$

The above results are in terms of LSTs; the joint distribution of the sample can in principle be evaluated by numerical inversion of the LSTs involved, but this is computationally demanding. Therefore, in the sequel we use a method-of-moments type estimator that requires finding the root of a function equating the empirical LST or the empirical moments to their respective expectations. The term $\psi(\xi) = \varphi^{-1}(\xi)$, which depends on the net input process as well as the sampling rate, appears in the above expressions. Estimating this quantity plays a crucial role in this work.

Notation. Much of this work involves large sample asymptotic analysis, and as such we will frequently make use of the following notations: $\to_{\mathbb{P}}$ for convergence in probability, $\to_{\mathrm{as}}$ for almost sure convergence, $\to_{\mathrm{d}}$ for convergence in distribution, and $\approx$ to indicate that two random sequences have the same limit in probability.

3. Semiparametric estimation of φ

This section proposes a Z-estimator for $\varphi(\alpha)$ for a given $\alpha > 0$. We then study the asymptotic properties of the estimator, in that we provide conditions for consistency and asymptotic normality given we know $\psi(\xi)$; the estimation of $\psi(\xi)$ is the topic of the next section. For brevity, we throughout denote $V_i := V(T_i)$.

Given a sample of workload observations at Poisson times, $(V_1, \ldots, V_n)$, let ψ_n denote a series of estimators for $\psi(\xi)$. In particular, ψ_n is a function $\mathbb{R}_+^n \to \mathbb{R}_+$ of the first n observations $(V_1, \ldots, V_n)$. We do not make any assumptions on ψ_n for now. Later, however, we impose conditions on this sequence that enable the evaluation of the asymptotic performance of the estimators, and in addition a specific construction for the special case of subordinator input.

An estimator for $\varphi(\alpha)$ is derived by equating the conditional LST (4) and the empirical LST and solving

$$\frac{1}{n}\sum_{i=1}^{n} e^{-\alpha V_i} = \frac{1}{n}\sum_{i=1}^{n} \frac{\xi}{\xi - \varphi(\alpha)}\left(e^{-\alpha V_{i-1}} - \frac{\alpha}{\psi_n}e^{-\psi_n V_{i-1}}\right). \tag{9}$$

Letting

$$Z_i(\alpha) := e^{-\alpha V_i} - \frac{\xi}{\xi - \varphi(\alpha)}\left(e^{-\alpha V_{i-1}} - \frac{\alpha}{\psi_n} e^{-\psi_n V_{i-1}}\right), \tag{10}$$

the estimator is given by the root of $\frac{1}{n}\sum_{i=1}^{n} Z_i(\alpha)$. We rearrange (9) to

$$\varphi(\alpha) = \xi\left[1 - \frac{1}{n}\sum_{i=1}^{n}\left(e^{-\alpha V_{i-1}} - \frac{\alpha}{\psi_n} e^{-\psi_n V_{i-1}}\right) \Big/ \frac{1}{n}\sum_{i=1}^{n} e^{-\alpha V_i}\right], \tag{11}$$

yielding the estimator

$$\hat{\varphi}_n(\alpha; \psi_n) = \frac{\xi\alpha}{\psi_n}\left(\frac{\psi_n}{\alpha n}\left(e^{-\alpha V_n} - e^{-\alpha V_0}\right) + \frac{1}{n}\sum_{i=1}^{n} e^{-\psi_n V_{i-1}}\right) \Big/ \frac{1}{n}\sum_{i=1}^{n} e^{-\alpha V_i}. \tag{12}$$

This type of estimator is generally referred to as a Z-estimator [35], Ch. V. Note that in our case the samples are dependent, thus preventing direct use of classical results on consistency and asymptotic normality. Another justification for this estimator is that it coincides with the estimator minimizing the sum of quadratic deviations of the empirical LST from their respective conditional expectations, that is, minimizing the conditional mean square error (MSE).

Remark 1. This method can also be applied for parametric estimation of the moments. For example, an estimator for the first moment of the net input process, that is, $\theta := \mathrm{E}X(1)$, can be derived. This is done by taking the derivative in (12):

$$\theta_n = -\hat{\varphi}_n'(0; \psi_n) = -\lim_{\alpha \downarrow 0} \frac{\mathrm{d}}{\mathrm{d}\alpha}\hat{\varphi}_n(\alpha; \psi_n) = -\frac{\xi}{n\psi_n}\sum_{i=1}^{n} e^{-\psi_n V_{i-1}}. \tag{13}$$

Alternatively, $\hat{\varphi}_n'(0; \psi_n)$ could be estimated directly from (6); then the estimation equation would also include the term $(V_n - V_0)/n$. Clearly, as n grows both estimators coincide. The same procedure can be followed for the (joint) estimation of higher moments. The estimators of the kth moment $\mathrm{E}(X(1)^k)$ follows from

$$\theta_n^{(k)} = (-1)^k \hat{\varphi}_n^{(k)}(0; \psi_n). \tag{14}$$

Remark 2. Recall that at $\alpha = \psi(\xi)$ the conditional LST $\mathrm{E}[e^{-\alpha V(T_i)} | V(T_{i-1})]$ is not given directly by (4) but can be computed using L'Hôpital's rule. In practice, this case will not play a role, though, as $\psi_n = \psi(\xi)$ will not occur in our setting in which various continuous random variables play a role. Specifically, while $\hat{\varphi}_n(\psi_n; \psi_n) = \xi$, by construction of the Z-estimator, the event $\hat{\varphi}_n(\psi(\xi); \psi_n) = \xi$ has zero probability to occur.

3.1. Consistency

The main result presented in this subsection is Theorem 1, claiming that consistency of the estimator ψ_n implies consistency of the estimator $\hat{\varphi}_n(\alpha; \psi_n)$ for all $\alpha > 0$. The proof relies on first

establishing the LLN for the empirical LST using the GPK formula (2) for the stationary distribution of the workload observations at Poisson sampling times. The result is then established by applying Slutsky's lemma and the continuous mapping theorem to the estimator given in (12).

Theorem 1. *If $\psi_n \to_{\mathbb{P}} \psi(\xi)$ as $n \to \infty$, then for every $\alpha > 0$ the pointwise estimator $\hat{\varphi}_n(\alpha; \psi_n)$ is consistent: $\hat{\varphi}_n(\alpha; \psi_n) \to_{\mathbb{P}} \varphi(\alpha)$ as $n \to \infty$.*

Proof. If $\mathrm{E}X(1) < 0$, then $(V(t))_{t\in\mathbb{R}}$ has a stationary distribution. By PASTA the limit distribution at Poisson sampling instants equals the stationary distribution of $(V(t))_{t\in\mathbb{R}}$, with corresponding random variable V. Denote the Poisson sampling process by $N(t)$, with $t_n := \inf\{t : N(t) = n - 1\}$. The choice of $n - 1$ in the definition of the counting process is made for the sake of brevity in the following analysis due to the observations always appearing as V_{i-1} in the estimation equation (9). Then, as $t_n \to \infty$ when $n \to \infty$, Equation (2) implies that for any $\beta \geq 0$,

$$\frac{1}{n}\sum_{i=1}^{n} e^{-\beta V_{i-1}} = \frac{1}{N(t_n)}\sum_{i=1}^{N(t_n)} e^{-\beta V(t_i)} \to_{\mathbb{P}} \mathrm{E}e^{-\beta V} = \frac{\beta\varphi'(0)}{\varphi(\beta)}, \tag{15}$$

as $n \to \infty$. We define the right-continuous process

$$H(t) := \sum_{n=1}^{\infty} \mathbf{1}(t_n \leq t < t_{n+1})\psi_n, \quad t \geq 0, \tag{16}$$

and observe that $H(t_n) = \psi_n$ for every $n \geq 1$. By Slutsky's lemma [35], p. 11, $\psi_n V(t_n) = H(t_n)V(t_n) \to_{\mathbb{P}} \psi(\xi)V$ as $n \to \infty$. Hence, applying PASTA once more,

$$\frac{1}{N(t_n)}\sum_{i=1}^{N(t_n)} e^{-H(t_n)V(t_i)} \to_{\mathbb{P}} \frac{\psi(\xi)\varphi'(0)}{\varphi(\psi(\xi))}. \tag{17}$$

Furthermore, the term $\frac{\psi_n}{\alpha n}(e^{-\alpha V_n} - e^{-\alpha V_0})$ converges to zero almost surely if the stability condition $\mathrm{E}X(1) < 0$ holds. Next we can use the fact that $\varphi(\psi(\xi)) = \varphi(\varphi^{-1}(\xi)) = \xi$ and apply the continuous mapping theorem (see [35], p. 7) to (12), thus obtaining

$$\hat{\varphi}_n(\alpha; \psi_n) \to_{\mathbb{P}} \left(\alpha\xi\frac{\varphi^{-1}(\xi)\varphi'(0)}{\varphi(\varphi^{-1}(\xi))}\right) \Big/ \left(\varphi^{-1}(\xi)\frac{\alpha\varphi'(0)}{\varphi(\alpha)}\right) = \varphi(\alpha). \tag{18}$$

Thus, the pointwise estimator of $\varphi(\alpha)$ is consistent for every $\alpha > 0$. □

Remark 3. Consider the case of parametric estimation of the (first and higher) moments, as described in Remark 1. Using the same arguments, it is straightforward to show that the estimators are consistent if ψ_n is consistent.

3.2. Asymptotic normality of $\hat{\varphi}_n$

In this section, we establish, under specific conditions, asymptotic normality of our estimator $\hat{\varphi}_n(\alpha;\psi_n)$ as $n\to\infty$: we show that $\sqrt{n}(\hat{\varphi}_n(\alpha;\psi_n)-\varphi(\alpha))$ converges to a zero-mean normal random variable, for any fixed $\alpha>0$. We find an expression for the corresponding asymptotic variance, so that one can assess the estimation error for large samples.

We then extend this result to a multivariate setting: we establish the joint normality of the vector

$$\sqrt{n}\big(\hat{\boldsymbol{\Phi}}_n(\boldsymbol{\alpha})-\Phi(\boldsymbol{\alpha})\big), \tag{19}$$

where $\boldsymbol{\alpha}:=(\alpha_1,\dots,\alpha_p)$, $\Phi(\boldsymbol{\alpha}):=(\varphi(\alpha_1),\dots,\varphi(\alpha_p))$ for some $p\in\mathbb{N}$, and $\hat{\boldsymbol{\Phi}}_n(\boldsymbol{\alpha}):=(\hat{\varphi}_n(\alpha_1;\psi_n),\dots,\hat{\varphi}_n(\alpha_p;\psi_n))$. Each of the estimation equations (9) is solved independently for α_i, but they rely on the same sample and are therefore dependent.

We start the section by stating a Central Limit Theorem (CLT) for martingale difference sums. Lemma 2 below consists of two parts; both are essentially known results, that we customized to our needs. The first part is the univariate result of [18], Thm. 3.2. The second part is a special case of the multidimensional extension of [25], Thm. 12.6; here we remark that the result in [25] relates to continuous time, but as we are only interested in the discrete-time case we present a more concise statement without the quadratic variation terms.

All convergence statements in the following are as $n\to\infty$; we therefore omit this specification from now on.

Lemma 2. (a) *Let M_n be a discrete-time martingale with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$, and denote*

$$Z_{ni}:=\frac{M_i-M_{i-1}}{\sqrt{\mathrm{Var}(M_n)}},\quad i=1,\dots,n. \tag{20}$$

If

$$\mathrm{E}\max_{1\leq i\leq n} Z_{ni}^2<\infty,\quad n=1,2,\dots \tag{21}$$

and

$$\max_{1\leq i\leq n}|Z_{ni}|\to_{\mathbb{P}}0, \tag{22}$$

then $\sum_{i=1}^n Z_{ni}\to_{\mathrm{d}}\mathrm{N}(0,1)$.

(b) *Let*

$$\boldsymbol{M}_n^\top:=\big(M_n^{(1)},\dots,M_n^{(p)}\big)$$

be a vector of discrete-time martingales, and let

$$\boldsymbol{Z}_n^\top:=\left(\sum_{i=1}^n Z_{ni}^{(1)},\dots,\sum_{i=1}^n Z_{ni}^{(p)}\right),$$

with $Z_{ni}^{(k)}$, *for* $k = 1, \dots, p$, *the respective normalized differences. If these objects satisfy all conditions of* (a), *and*

$$\lim_{n\to\infty} \mathrm{E}\big[\boldsymbol{Z}_n \boldsymbol{Z}_n^\top\big] = S, \tag{23}$$

where $S \in \mathbb{R}^{p\times p}$ *is a positive-definite matrix with finite elements, then* $\boldsymbol{Z}_n \to_{\mathrm{d}} \mathrm{N}(0, S)$.

3.2.1. *Univariate asymptotic normality*

Recall that the estimation procedure also relies on an external estimator for $\psi(\xi)$. As the CLT below shows, the asymptotic distribution of $\sqrt{n}(\psi_n - \psi(\xi))$ and its joint distribution with the estimator for $\varphi(\alpha)$ play a crucial role in the analysis.

Theorem 3. *Suppose that* $\mathrm{E}X(1) < 0$ *and* $\psi_n - \psi(\xi) \approx \frac{1}{n} M_n + R_n$ *such that*:

(i) $\sqrt{n} R_n \to_{\mathbb{P}} 0$ *and* M_n *is a martingale with respect to* $(V_0, \dots, V_n)$ *that satisfies the conditions of Lemma* 2(*a*),
(ii) $\lim_{n\to\infty} \mathrm{Var}(M_n)/n = \sigma_\xi^2 < \infty$,
(iii) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \mathrm{E}[Z_i(\alpha) M_n] < \infty$,

where $Z_i(\alpha)$ *is given by* (10). *Then* $\sqrt{n}(\psi_n - \psi(\xi)) \to_{\mathrm{d}} \mathrm{N}(0, \sigma_\xi^2)$, *and for every* $\alpha > 0$,

$$\sqrt{n}\big(\hat{\varphi}_n(\alpha; \psi_n) - \varphi(\alpha)\big) \to_{\mathrm{d}} \mathrm{N}\big(0, \sigma_{\alpha,\xi}^2\big), \tag{24}$$

where $0 < \sigma_{\alpha,\xi}^2 < \infty$.

Remark 4. The normality assumptions involving the asymptotic estimation error of $\psi(\xi)$ have to be verified for the specific ψ_n used. Condition (i) entails that the estimation error satisfies a martingale CLT, condition (ii) specifies the convergence rate of the variance and can in principle be generalized to other rates, and condition (iii) further demands that the asymptotic covariance between estimation errors is bounded with the appropriate scaling. As it turns out, under some regularity conditions these assumptions are valid for various estimation procedures (see, e.g., [25]). In the next section we present an MLE for $\psi(\xi)$ for the subordinator case, and show that it satisfies the conditions of Theorem 3.

Proof of Theorem 3. In this proof, a number of lemmas are needed. These are stated after the end of the proof, and are proven in the supplementary material [31].

We focus on the marginal asymptotic distribution of the estimation error for a fixed $\alpha > 0$. We denote the estimator by $\varphi_n := \hat{\varphi}_n(\alpha; \psi_n)$ and recall that the true value at α is $\varphi(\alpha)$.

The starting point is the estimation equation (9), which can be written as

$$J_n(\psi, \varphi) := \frac{1}{n} \sum_{i=1}^n \left[e^{-\alpha V_i} - \frac{\xi}{\xi - \varphi} \left(e^{-\alpha V_{i-1}} - \frac{\alpha}{\psi} e^{-\psi V_{i-1}} \right) \right] = 0. \tag{25}$$

The standard *delta method* can be applied for deriving the asymptotic distribution of φ_n; see, for example, [35], p. 51. Concretely, taking a Taylor expansion of $J_n(\psi_n, \varphi_n)$, where ψ_n and φ_n are

the estimators, around the true parameters $\psi(\xi)$ and $\varphi(\alpha)$, yields

$$J_n(\psi_n,\varphi_n) \approx J_n\big(\psi(\xi),\varphi(\alpha)\big) + \big(\psi_n-\psi(\xi)\big)\frac{\partial}{\partial\psi}J_n\big(\psi(\xi),\varphi(\alpha)\big) + \big(\varphi_n-\varphi(\alpha)\big)\frac{\partial}{\partial\varphi}J_n\big(\psi(\xi),\varphi(\alpha)\big), \tag{26}$$

where the remainder will be argued to vanish, in a convergence-in-probability sense, in Lemma 6. As $J_n(\psi_n,\varphi_n)=0$, we thus obtain

$$\sqrt{n}\big(\varphi_n-\varphi(\alpha)\big) \approx \frac{\sqrt{n}J_n(\psi(\xi),\varphi(\alpha)) + \sqrt{n}(\psi_n-\psi(\xi))\frac{\partial}{\partial\psi}J_n(\psi(\xi),\varphi(\alpha))}{-\frac{\partial}{\partial\varphi}J_n(\psi(\xi),\varphi(\alpha))}. \tag{27}$$

Lemma 4 states that partial derivatives featuring in the right-hand side of (27) converge almost surely to constants, which we call ∂J_ψ and ∂J_φ. Thus, by applying Slutsky's lemma to (27), $\sqrt{n}(\varphi_n-\varphi(\alpha))$ converges in distribution to the same limit as

$$-\frac{1}{\partial J_\varphi}\big(\sqrt{n}J_n\big(\psi(\xi),\varphi(\alpha)\big) + \sqrt{n}\big(\psi_n-\psi(\xi)\big)\partial J_\psi\big). \tag{28}$$

By condition (i),

$$\sqrt{n}\big(\psi_n-\psi(\xi)\big) \approx \frac{1}{\sqrt{n}}M_n = \sqrt{\frac{\mathrm{Var}(M_n)}{n}}\sum_{i=1}^{n} Z_{ni},$$

where $\sum_{i=1}^{n} Z_{ni}$ is a normalized sum of martingale differences that, due to Lemma 2(a), converges to a standard normal random variable. We conclude, by condition (ii), that the limiting distribution of $\sqrt{n}(\psi_n-\psi(\xi))$ is normal. In Lemma 5a, we further show that $\sqrt{n}J_n(\psi(\xi),\varphi(\alpha)) \to_{\mathrm{d}} \mathrm{N}(0,\sigma_\alpha^2)$, where σ_α^2 is given by (34). In order to complete the proof of Theorem 3, it is now sufficient to prove that the limiting joint distribution of $\sqrt{n}J_n(\psi(\xi),\varphi(\alpha))$ and $\sqrt{n}(\psi_n-\psi(\xi))$ is (bivariate) normal, or equivalently that any linear combination of them is asymptotically normal. This is done in Lemma 5b by applying Lemma 2(b), which is facilitated by the assumption of condition (iii). Finally, we conclude that the asymptotic variance of $\sqrt{n}(\varphi_n-\varphi(\alpha))$ is

$$\sigma_{\alpha,\xi}^2 := \frac{\sigma_\alpha^2 + 2\partial J_\psi\sigma_{\alpha,\psi}^2 + (\partial J_\psi)^2\sigma_\xi^2}{(\partial J_\varphi)^2}, \tag{29}$$

where

$$\sigma_{\alpha,\psi}^2 := \lim_{n\to\infty}\mathrm{Cov}\big[\sqrt{n}J_n\big(\psi(\xi),\varphi(\alpha)\big), \sqrt{n}\big(\psi_n-\psi(\xi)\big)\big]. \tag{30}$$

This proves the claimed asymptotic normality. □

Remark 5. The explicit form of the term $\sigma_{\alpha,\psi}^2$ depends on the specific estimator ψ_n. In the sequel, we derive an explicit expression for the asymptotic variance in the subordinator case when ψ_n is an MLE.

Lemma 4. *As $n \to \infty$,*

$$\frac{\partial}{\partial \psi} J_n\big(\psi(\xi), \varphi(\alpha)\big) \to_{\mathrm{as}} -\frac{\alpha \varphi'(0)}{(\xi - \varphi(\alpha))\psi'(\xi)} =: \partial J_\psi \tag{31}$$

and

$$\frac{\partial}{\partial \varphi} J_n\big(\psi(\xi), \varphi(\alpha)\big) \to_{\mathrm{as}} -\frac{\alpha \varphi'(0)}{\varphi(\alpha)(\xi - \varphi(\alpha))} =: \partial J_\varphi. \tag{32}$$

Lemma 5. (a) *For any $\alpha > 0$,*

$$\sqrt{n} J_n\big(\psi(\xi), \varphi(\alpha)\big) \to_{\mathrm{d}} \mathrm{N}\big(0, \sigma_\alpha^2\big), \tag{33}$$

where

$$\sigma_\alpha^2 := \frac{2\xi^2 \alpha \varphi'(0)}{(\xi - \varphi(\alpha))^2} \left(\frac{(\frac{\varphi(\alpha)}{\xi})^2 - \frac{2\varphi(\alpha)}{\xi}}{\varphi(2\alpha)} - \frac{\alpha}{\psi(\xi)\varphi(2\psi(\xi))} + \frac{\alpha + \psi(\xi)}{\psi(\xi)\varphi(\alpha + \psi(\xi))} \right). \tag{34}$$

(b) *Furthermore, under the conditions of Theorem 3,*

$$\sqrt{n} \begin{pmatrix} J_n\big(\psi(\xi), \varphi(\alpha)\big) \\ \psi_n - \psi(\xi) \end{pmatrix} \to_{\mathrm{d}} \mathrm{N}\left(\mathbf{0}, \begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha,\psi}^2 \\ \sigma_{\alpha,\psi}^2 & \sigma_\xi^2 \end{pmatrix} \right). \tag{35}$$

Consequently,

$$\sqrt{n}\big(J_n\big(\psi(\xi), \varphi(\alpha)\big) + \partial J_\psi\big(\psi_n - \psi(\xi)\big)\big) \to_{\mathrm{d}} \mathrm{N}\big(0, \sigma_\alpha^2 + 2\partial J_\psi \sigma_{\alpha,\psi}^2 + (\partial J_\psi)^2 \sigma_\xi^2\big). \tag{36}$$

Lemma 6. *The remainder term of the Taylor expansion in* (26) *converges in probability to zero.*

3.2.2. *Multivariate asymptotic normality*

We now consider the situation that $\varphi(\alpha)$ is estimated for multiple values of α simultaneously. The main result is that the estimation error of $\hat{\mathbf{\Phi}}_n(\boldsymbol{\alpha})$ converges to a multivariate normal random variable at rate $\sqrt{n}$, under the conditions of Theorem 3. The proof combines the lines of reasoning developed for the univariate asymptotic normality with the use of the Cramér–Wold device.

We first introduce some notation. Denote for any $x > 0$,

$$\partial J_{\psi,x} := \lim_{n\to\infty} \frac{\partial}{\partial \psi} J_n\big(\psi(\xi), \varphi(x)\big), \qquad \partial J_{\varphi,x} := \lim_{n\to\infty} \frac{\partial}{\partial \varphi} J_n\big(\psi(\xi), \varphi(x)\big), \tag{37}$$

as defined in (31) and (32). In addition, let

$$\sigma_{J,\alpha,\beta}^2 := \frac{(\alpha+\beta)\varphi'(0)}{\varphi(\alpha+\beta)} - \frac{\xi^2 \varphi'(0)}{(\xi - \varphi(\alpha))(\xi - \varphi(\beta))} \times \left(\frac{\alpha+\beta}{\varphi(\alpha+\beta)} + \frac{2\alpha\beta}{\psi(\xi)\varphi(2\psi(\xi))} - \frac{\beta(\alpha + \psi(\xi))}{\psi(\xi)\varphi(\alpha + \psi(\xi))} \right.$$

$$-\frac{\alpha(\beta+\psi(\xi))}{\psi(\xi)\varphi(\beta+\psi(\xi))}\Bigg), \tag{38}$$

and, for every $\alpha \neq \beta$,

$$\sigma^2_{\alpha,\beta} := \frac{\sigma^2_{J,\alpha,\beta} + \partial J_{\psi,\beta}\sigma^2_{\alpha,\xi} + \partial J_{\psi,\alpha}\sigma^2_{\beta,\xi} + \partial J_{\psi,\alpha}\partial J_{\psi,\beta}\sigma^2_{\xi}}{\partial J_{\varphi,\alpha}\partial J_{\varphi,\beta}}. \tag{39}$$

Theorem 7. *Suppose that* $\mathrm{E}X(1) < 0$. *Let* $\Phi(\boldsymbol{\alpha}) := (\varphi(\alpha_1), \dots, \varphi(\alpha_p))$ *and* $\hat{\boldsymbol{\Phi}}_n(\boldsymbol{\alpha}) := (\hat{\varphi}_n(\alpha_1; \psi_n), \dots, \hat{\varphi}_n(\alpha_p; \psi_n))$, *where* $(\alpha_1, \dots, \alpha_p) \in (0, \infty)^p$. *If* ψ_n *satisfies the conditions of Theorem* 3, *and in particular* (*iii*) *is satisfied for every* α_i, $i = 1, \dots, p$, *then*

$$\sqrt{n}\big(\hat{\boldsymbol{\Phi}}_n(\boldsymbol{\alpha}) - \Phi(\boldsymbol{\alpha})\big) \to_{\mathrm{d}} \mathrm{N}(0, \Sigma), \tag{40}$$

where $\Sigma_{ij} = \sigma^2_{\alpha_i,\xi}$ *if* $i = j$ *and* $\sigma^2_{\alpha_i,\alpha_j}$ *otherwise.*

4. Estimating $\psi(\xi)$ for subordinator input

Theorems 1 and 3 have established that, under specific conditions, the Z-estimator for $\varphi(\alpha)$ is consistent (asymptotically normal, respectively) if we have a consistent (asymptotically normal) estimator for $\psi(\xi)$. In this section, we propose such an estimator for $\psi(\xi)$ using a likelihood approach for the subordinator input case, that is, the input process has almost surely non-decreasing paths. An important special case is the CP model: nonnegative jumps (with LST $G^*(\alpha)$) arrive according to a Poisson process with rate λ. We thus have that for this example,

$$\varphi(\alpha) = \lambda\big(G^*(\alpha) - 1\big) + \alpha. \tag{41}$$

In the following an ML-type estimator is constructed for $\psi(\xi)$. Suppose we have a sample of workload observations $\boldsymbol{V} = (V_0, V_1, \dots, V_n)$. Let $Y_i := \mathbf{1}(V_i = 0)$, then by (8),

$$\mathrm{P}(Y_i = 1|\boldsymbol{V}) = \frac{\xi}{\psi(\xi)} e^{-\psi(\xi)V_{i-1}}, \quad i = 1, \dots, n. \tag{42}$$

Conditional on $\mathbf{V}$, the sample $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ is then distributed as a sample of independent, but non-identically distributed, Bernoulli trials, with likelihood

$$\mathrm{P}_{\psi(\xi)}(Y_1, \dots, Y_n|\mathbf{V}) = \prod_{i=1}^{n}\left[\frac{\xi}{\psi(\xi)} e^{-\psi(\xi)V_{i-1}}\right]^{Y_i}\left[1 - \frac{\xi}{\psi(\xi)} e^{-\psi(\xi)V_{i-1}}\right]^{1-Y_i}. \tag{43}$$

We can thus define the MLE for $\psi(\xi)$: $\hat{\psi}_n = \arg\max_{\psi\in\Psi}\{\mathrm{P}_\psi(Y_1, \dots, Y_n|\mathbf{V})\}$.

We assume that the parameter space Ψ is compact and that the true value lies within the interior. This assumption is reasonable and can be achieved by setting $\Psi = [\xi, C]$, where C is an arbitrarily large constant. Note that $\psi(\xi) \geq \xi$ by (8). For special cases, the parameter space

can be specified in a more detailed manner. For example, the exponent function for the CP case is given in (41), and as $G^*(\alpha) \in [0, 1]$, it follows that $\alpha - \lambda \le \varphi(\alpha) \le \alpha$. This implies

$$\psi(\xi) - \lambda \le \varphi\big(\psi(\xi)\big) \le \psi(\xi). \tag{44}$$

As $\psi(\xi) = \varphi^{-1}(\xi)$, we conclude that the parameter space is $\Psi = [\xi, \xi + \lambda]$. In this definition there is an implicit assumption that λ is known. If it is unknown, one could again follow the pragmatic approach of replacing $\xi + \lambda$ by an arbitrarily large constant.

For any $\psi \in \Psi$, the log-likelihood function, $L_n(\psi) = \log \mathrm{P}_\psi(Y_1, \dots, Y_n | \boldsymbol{V})$, is

$$L_n(\psi) = \sum_{i=1}^{n} \left[Y_i(\log \xi - \log \psi - \psi V_{i-1}) + (1 - Y_i) \log\left(1 - \frac{\xi}{\psi} e^{-\psi V_{i-1}}\right)\right], \tag{45}$$

and taking derivatives yields

$$L_n'(\psi) = \sum_{i=1}^{n} \frac{(\frac{1}{\psi} + V_{i-1})(\frac{\xi}{\psi} e^{-\psi V_{i-1}} - Y_i)}{1 - \frac{\xi}{\psi} e^{-\psi V_{i-1}}}. \tag{46}$$

The MLE is the solution of the nonlinear optimization problem,

$$\hat{\psi}_n := \arg\max_{\psi \in \Psi} L_n(\psi). \tag{47}$$

The function $L_n(\psi)$ is not always concave with respect to ψ but as the parameter space is compact a maximizer always exists. If $Y_i = 1$ for all $i = 1, \dots, n$ (i.e., all observations are at idle periods), then by (43) the likelihood $L_n(\psi)$ decreases with ψ, and therefore the MLE is the lower boundary of the parameter space. On the other hand if $Y_i = 0$ for all $i = 1, \dots, n$ (i.e., no idle periods are observed), then by (43) the likelihood $L_n(\psi)$ increases with ψ, and then the MLE is the upper boundary of the support. Otherwise, the solution may be in the interior or boundary of the parameter space. In the following subsection, we will show that as n increases the asymptotic log-likelihood has a unique maximizer, that solves the first order equation $L_n'(\hat{\psi}_n) = 0$.

4.1. Consistency

In this subsection, we establish strong consistency of the MLE $\hat{\psi}_n$. As $\hat{\psi}_n$ is consistent (actually even strongly consistent), Theorem 1 thus yields the consistency of $\hat{\varphi}_n(\alpha; \hat{\psi}_n)$. The key for determining consistency and asymptotic normality is the asymptotic analysis of the log-likelihood function given in (45) and its first two derivatives. We define $\ell_n(\psi) := n^{-1} L_n(\psi)$, and $\ell_n'(\psi)$ and $\ell_n''(\psi)$ as the respective first two derivatives. There are several issues that prevent us from directly applying standard MLE consistency results. Firstly, the observations are dependent. Additionally, the function $|\ell_n(\psi)|$ may be unbounded at the boundaries of the parameter space: for example, if there is some i such that $Y_i = 1$ and $V_{i-1} = 0$ then $\ell_n(\psi) \to -\infty$ as $\psi \downarrow \xi$. Therefore, we will need to verify that as n grows the MLE moves away from the boundary almost surely.

In the next theorem, we assume that Ψ is compact and that the true parameter belongs to the interior of the parameter space, Ψ^{o}. In practice, setting $\Psi = [\xi, C]$ for a very large C should ensure this condition.

Theorem 8. *If* $\mathrm{E}X(1) < 0$, *and if the parameter space* Ψ *is compact and* $\psi(\xi) \in \Psi^{\mathrm{o}}$, *then the MLE is strongly consistent*: $\hat{\psi}_n \to_{\mathrm{as}} \psi(\xi)$.

Upon combining Theorems 1 and 8, we thus have the following result.

Corollary 9. *If* $\mathrm{E}X(1) < 0$, *and if the parameter space* Ψ *is compact and* $\psi(\xi) \in \Psi^{\mathrm{o}}$, *then the Z-estimator for* $\varphi(\alpha)$ *defined in* (12) *using the MLE for* $\psi(\xi)$ *is pointwise consistent*: $\hat{\varphi}_n(\alpha; \hat{\psi}_n) \to_{\mathbb{P}} \varphi(\alpha)$, *for all* $\alpha > 0$.

The remainder of this subsection focuses on proving Theorem 8. We commence by providing two lemmas, which are proven in the supplementary material [31], featuring properties of the log-likelihood function that will be utilized in the asymptotic analysis. In particular, we assert smoothness and the construction of a strong law of large numbers for the first two derivatives of the log-likelihood.

Lemma 10. *If the parameter space* $\Psi \subset [\xi, \infty)$ *is a compact interval, then the functions* $\ell_n(\psi)$ *and* $\ell_n'(\psi)$ *are bounded and smooth on* Ψ^{o}.

Lemma 11. *If* $\mathrm{E}X(1) < 0$, *then for all* $\psi \in \Psi^{\mathrm{o}}$,

$$\ell_n'(\psi) \to_{\mathrm{as}} \mathrm{E}\big(\mathrm{E}\big(\ell_1'(\psi)|V\big)\big) = \mathrm{E}\left[\frac{\frac{1}{\psi} + V}{1 - \frac{\xi}{\psi} e^{-\psi V}} \left(\frac{\xi}{\psi} e^{-\psi V} - \frac{\xi}{\psi(\xi)} e^{-\psi(\xi) V}\right)\right], \tag{48}$$

and

$$\ell_n''(\psi) \to_{\mathrm{as}} \mathrm{E}\big(\mathrm{E}\big(\ell_1''(\psi)|V\big)\big). \tag{49}$$

Proof of Theorem 8. The following condition is sufficient for the consistency of an MLE with dependent observations [23]: for every $\psi \neq \psi(\xi)$ (in the interior of Ψ) there exists a $\delta > 0$ such that

$$\lim_{n \to \infty} \sup_{\zeta \in \mathcal{B}_\delta(\psi)} \big\{\ell_n(\zeta) - \ell_n\big(\psi(\xi)\big)\big\} < 0, \tag{50}$$

almost surely, where $\mathcal{B}_\delta(\psi) := (\psi - \delta, \psi + \delta)$ is an open ball around ψ. Firstly, applying (48) yields

$$\ell_n'(\psi) \to_{\mathrm{as}} \mathrm{E}\big(\mathrm{E}\big(\ell_1'(\psi)|V\big)\big) \begin{cases} > 0, & \psi < \psi(\xi), \\ = 0, & \psi = \psi(\xi), \\ < 0, & \psi > \psi(\xi). \end{cases} \tag{51}$$

As $\psi(\xi) \in \Psi^{\mathrm{o}}$, for $\psi < \psi(\xi)$ we can consider a ball such that $\psi + \delta < \psi(\xi)$, and then for any $\zeta < \psi + \delta < \psi(\xi)$, by Lemma 10 the difference can be written as

$$\ell_n(\zeta) - \ell_n\big(\psi(\xi)\big) = -\big(\ell_n\big(\psi(\xi)\big) - \ell_n(\zeta)\big) = -\int_{\zeta}^{\psi(\xi)} \ell_n'(x)\,\mathrm{d}x. \tag{52}$$

By (51), we have that as $n \to \infty$, $\ell_n'(\psi) > 0$ almost surely for any $\psi \in [\psi + \delta, \psi(\xi)]$, which is a non-empty interval, and thus

$$\sup_{\zeta \in (\psi-\delta, \psi+\delta)} \big\{\ell_n(\zeta) - \ell_n\big(\psi(\xi)\big)\big\} \le -\int_{\psi+\delta}^{\psi(\xi)} \ell_n'(x)\,\mathrm{d}x < 0 \tag{53}$$

almost surely. Similarly, for any $\psi > \psi(\xi)$ we can consider a ball such that $\psi - \delta > \psi(\xi)$ and conclude that

$$\sup_{\zeta \in (\psi-\delta, \psi+\delta)} \big\{\ell_n(\zeta) - \ell_n\big(\psi(\xi)\big)\big\} \le \int_{\psi(\xi)}^{\psi-\delta} \ell_n'(x)\,\mathrm{d}x < 0 \tag{54}$$

almost surely. Therefore, for any $\psi \neq \psi(\xi)$ there exists a $\delta < |\psi - \psi(\xi)|$ such that (50) holds, almost surely. As a consequence, (50) is satisfied. □

4.2. Asymptotic normality

We now discuss the asymptotic distribution of the MLE $\hat{\psi}_n$. In this subsection, we show that the estimation term is asymptotically normal with rate $\sqrt{n}$, and that it further satisfies the conditions of Theorem 3. Useful conditions for asymptotic normality of the MLE when the observations are dependent are given in [24]. Again the delta method is relied on; we also make use of arguments used in the proof of Cramér's theorem [13], p. 121. Define

$$I_\xi := \mathrm{E}\left[\frac{\frac{\xi}{\psi(\xi)} e^{-\psi(\xi)V} (\frac{1}{\psi(\xi)} + V)^2}{1 - \frac{\xi}{\psi(\xi)} e^{-\psi(\xi)V}}\right]. \tag{55}$$

Theorem 12. *If* $\mathrm{E}X(1) < 0$, *and if the parameter space* Ψ *is compact and* $\psi(\xi) \in \Psi^{\mathrm{o}}$, *then*

$$\sqrt{n}\big(\hat{\psi}_n - \psi(\xi)\big) \to_{\mathrm{d}} \mathrm{N}\left(0, \frac{1}{I_\xi}\right). \tag{56}$$

We first outline the two main steps required to establish the asymptotic distribution. This will be followed by a separate lemma verifying each of these steps.

Proof. The smoothness of $\ell_n'(\cdot)$ established in Lemma 10 implies

$$\ell_n'(\hat{\psi}_n) = \ell_n'\big(\psi(\xi)\big) + \int_{\psi(\xi)}^{\hat{\psi}_n} \ell_n''(x)\,\mathrm{d}x, \tag{57}$$

and by applying a change of variables

$$\ell_n'(\hat{\psi}_n) = \ell_n'\big(\psi(\xi)\big) + \big(\hat{\psi}_n - \psi(\xi)\big)\int_0^1 \ell_n''\big(\psi(\xi) + y\big(\hat{\psi}_n - \psi(\xi)\big)\big)\,\mathrm{d}y$$
$$= \ell_n'\big(\psi(\xi)\big) + \big(\hat{\psi}_n - \psi(\xi)\big)\int_0^1 \ell_n''\big(y\hat{\psi}_n + (1-y)\psi(\xi)\big)\,\mathrm{d}y. \tag{58}$$

If $\psi(\xi)$ is in the interior of Ψ, then by Lemma 11, as $n \to \infty$, the MLE is given by the root of the first order condition $L_n'(\hat{\psi}_n) = 0$, almost surely. To verify this conclusion one only has to consider the limits of $\ell_n'(\psi)$ at the lower and upper boundary points of the compact interval Ψ, which are almost surely positive and negative, respectively. Therefore, $\sqrt{n}\ell_n'(\hat{\psi}_n) = \frac{1}{\sqrt{n}}L_n'(\hat{\psi}_n) \to_{\mathrm{as}} 0$. Hence, denoting

$$D_n := \int_0^1 \ell_n''\big(y\hat{\psi}_n + (1-y)\psi(\xi)\big)\,\mathrm{d}y, \tag{59}$$

we have

$$\sqrt{n}\ell_n'\big(\psi(\xi)\big) + \sqrt{n}\big(\hat{\psi}_n - \psi(\xi)\big)D_n \approx 0, \tag{60}$$

and therefore

$$\sqrt{n}\big(\hat{\psi}_n - \psi(\xi)\big) \approx \frac{\sqrt{n}\ell_n'(\psi(\xi))}{-D_n}. \tag{61}$$

We show that $\sqrt{n}\ell_n'(\psi(\xi)) \to_{\mathrm{d}} N(0, I_\xi)$ and $D_n \to_{\mathrm{as}} -I_\xi$ in Lemmas 13 and 14, respectively. Then, by Slutsky's lemma, (61) implies (56). □

Lemma 13. *Under the conditions of Theorem 8,* $\sqrt{n}\ell_n'(\psi(\xi)) \to_{\mathrm{d}} \mathrm{N}(0, I_\xi)$.

Lemma 14. *Under the conditions of Theorem 8,* $D_n \to_{\mathrm{as}} -I_\xi$.

For the proofs of these two lemmas we refer to the supplementary material [31].

Remark 6. The asymptotic variance (55) in Theorem 12 is $\sigma_\xi^2 := 1/I_\xi$. In the CP case $\varphi(\alpha) = \lambda(G^*(\alpha) - 1) + \alpha$, and therefore $\psi(\xi)$ is an increasing function of ξ such that $\psi(0) = 0$ and

$$\frac{\xi}{\psi(\xi)} = \frac{\lambda(G^*(\psi(\xi)) - 1) + \psi(\xi)}{\psi(\xi)} = 1 + \lambda\frac{G^*(\psi(\xi)) - 1}{\psi(\xi)} \xrightarrow{\xi \downarrow 0} \varphi'(0). \tag{62}$$

This implies that the asymptotic variance σ_ξ^2 increases with the sampling rate ξ, and that for a decreasing rate the variance disappears: $\sigma_\xi^2 \to 0$ as $\xi \downarrow 0$.

4.3. Asymptotic error of $\hat{\varphi}_n(\alpha; \hat{\psi}_n)$

Theorem 12 has implications also for the asymptotic estimation error of $\varphi(\alpha)$. We verified the martingale CLT conditions of Lemma 2(a) for $\sqrt{n}(\psi_n - \psi(\xi))$ in Theorem 12, and thus if we fur-

ther show that the covariance of the estimation errors has a finite limit we can apply Lemma 2(b) and Theorem 3 to conclude that $\sqrt{n}(\varphi_n(\alpha;\hat{\psi}_n)-\varphi(\alpha))$ is asymptotically normal. This is done in the proof of Theorem 15 in the supplementary material [31].

Theorem 15. *If* $\mathrm{E}X(1)<0$, $\hat{\psi}_n$ *is the MLE for* $\psi(\xi)$, *then*

$$\sqrt{n}\big(\varphi_n(\alpha;\hat{\psi}_n)-\varphi(\alpha)\big)\to_{\mathrm{d}}\mathrm{N}\big(0,\sigma^2_{\alpha,\xi}\big), \tag{63}$$

where $\sigma^2_{\alpha,\xi}$ *is given by* (29) *with*

$$\sigma^2_{\alpha,\psi}:=\frac{\xi^2}{\psi(\xi)(\xi-\varphi(\alpha))I_\xi}\times\mathrm{E}\left[\frac{(\frac{1}{\psi(\xi)}+V)e^{-\psi(\xi)V}}{1-\frac{\xi}{\psi(\xi)}e^{-\psi(\xi)V}}\left(e^{-\alpha V}-\frac{\alpha}{\psi(\xi)}e^{-\psi(\xi)V}-\frac{\xi-\varphi(\alpha)}{\xi}\right)\right]. \tag{64}$$

5. Estimating $\psi(\xi)$ for spectrally-positive Lévy input

The Z-estimator $\hat{\varphi}_n(\alpha;\psi_n)$ requires an estimator for $\psi(\xi)$. In Section 4, we considered the subordinator case, utilizing that there is a non-zero probability that the queue is empty. For general spectrally-positive Lévy input with a Brownian motion component this property does not hold. In this section, we present an alternative procedure.

A moment based Z-estimator ψ_n can be derived using (6) by solving

$$\sum_{i=1}^n(V_i-V_{i-1})-\sum_{i=1}^n\left(\frac{e^{-\psi_n V_{i-1}}}{\psi_n}-\frac{\varphi'(0)}{\xi}\right)=0. \tag{65}$$

After some rearranging of the terms, we define the estimator ψ_n as the solution of

$$V_n-V_0+\frac{n\varphi'(0)}{\xi}=\sum_{i=1}^n\frac{e^{-\psi_n V_{i-1}}}{\psi_n}, \tag{66}$$

when it exists. There is at most one solution as the left-hand side is constant and the right-hand side is decreasing from infinity to zero on $\psi_n\in\mathbb{R}_+$. However, if the constant on the left is negative there is no solution and then we can set $\psi_n=\bar{\psi}$, where $\bar{\psi}>0$ is some arbitrary large constant. Observe that under the stability condition $\varphi'(0)=-\mathrm{E}X(1)>0$, the probability of having a negative term on the left vanishes as $n\to\infty$.

The problem, however, is that $\varphi'(0)$ in (66) is also unknown and needs to be somehow estimated. Note that if we naïvely plug $\hat{\varphi}'_n(0;\psi_n)$ into (66) then all terms including ψ_n cancel out because the estimator of the first derivative relies on the very same equation, that is, (6). In the next subsections, we will present an estimation scheme for $\varphi'(0)$ that relies on considering a sub-sample with high initial workloads above some given threshold. We also show how the estimator's inherent bias can be made arbitrarily small.

5.1. Estimating $\varphi'(0)$

We construct an estimator for $\varphi'(0)$ relying on the fact that $\xi \mathrm{E}X(T) = -\varphi'(0)$ where $T \sim \exp(\xi)$, and estimating the net input only using observations that are far from zero. If a high workload was observed then the probability of reflection (i.e., the queue hitting zero) before the next observation is low. In particular, we can quantify this probability so as to provide an upper bound for the bias of the estimator.

For some (large) threshold $\tau > 0$ the workload is sampled until there are $m \in \mathbb{N}$ observations V_i such that $V_{i-1} \geq \tau$. Let $M(m, \tau)$ denote the total number of observations, which is now a random variable, and denote the jth such observation by $i(j) \in \{1, \ldots, M(m, \tau)\}$, for $j = 1, \ldots, m$. The estimator for $\vartheta = -\varphi'(0) = \xi \mathrm{E}X(T)$ is

$$\hat{\vartheta}_m(\tau) := \frac{\xi}{m} \sum_{j=1}^{m} (V_{i(j)} - V_{i(j)-1}). \tag{67}$$

Observe that the statistic $\hat{\vartheta}_m(\tau)$ only depends on the increments of the workload process which makes it convenient for analysis due to the independent increments property of Lévy processes. We first derive an upper bound for the bias of $\hat{\vartheta}_m(\tau)$ and in the sequel provide an accurate asymptotic analysis of the bias. The proofs for this section are detailed in the supplementary material [31].

Proposition 16. *Let $b_m(\hat{\vartheta}; \tau) := \mathrm{E}[\hat{\vartheta}_m(\tau) - \vartheta]$. Then, for any $m \in \mathbb{N}$ and $\tau > 0$,*

$$0 < b_m(\hat{\vartheta}; \tau) \leq \xi \mathrm{E}\big[V(T) | V(0) = 0\big] e^{-\psi(\xi)\tau}. \tag{68}$$

The bound (68) for the bias decays to zero as τ grows, but at the same time the probability of observing workloads above τ also diminishes if $\mathrm{E}X(1) < 0$, so that the required number of samples $M(m, \tau)$ also grows. This entails that there is a tradeoff between τ and m in terms of accuracy, information and required sample size. The expectation in the bound, i.e., $\mathrm{E}[V(T)|V(0) = 0]$, can be approximately estimated by using observations $V_i - V_{i-1}$ such that $V_{i-1} \leq \epsilon$ for some low threshold $\epsilon > 0$. The bound (68) also depends on $\psi(\xi)$ which can be estimated using (66) by plugging in the estimator $\hat{\vartheta}_m(\tau)$ instead of $-\varphi'(0)$. There is a (seemingly) circular argument here: the estimated bias bound is also biased because it also relies on the biased estimator $\hat{\vartheta}_m$. In the next proposition, however, we argue that this bias is bounded as well and diminishes with τ.

Proposition 17. *Let $\tilde{\psi}_m$ be the Z-estimator given by (66) with $\varphi'(0)$ replaced with the estimator $\hat{\vartheta}_m(\tau)$, then the bias $\mathrm{E}[\tilde{\psi}_m - \psi(\xi)]$ diminishes to zero as $\tau \to \infty$.*

5.2. Asymptotic analysis

In this subsection, we consider asymptotic properties of $\hat{\vartheta}_m(\tau)$ as $m \to \infty$. For any $\tau < \infty$, there is positive bias due to the possibility of reflection, even though it may be very small, and

therefore the estimator cannot be consistent. A possible solution to this is defining an estimator with a dynamic threshold τ_m that grows with the sample size. In this section we analyze the asymptotic properties of the biased estimator based on the static threshold τ. We first establish that the estimator $\hat{\vartheta}_m(\tau)$ converges in probability (as $m \to \infty$, that is) to a constant that depends on τ. This further enables the computation of the exact asymptotic bias.

Proposition 18. *If* $\mathrm{E}X(1) < 0$ *then for any* $\tau > 0$, *as* $m \to \infty$, $\hat{\vartheta}_m(\tau) \to_{\mathbb{P}} \vartheta(\tau)$, *where*

$$\vartheta(\tau) := \frac{\xi \mathrm{E}[e^{-\psi(\xi)V}\mathbf{1}(V \geq \tau)]}{\psi(\xi)\mathrm{P}(V \geq \tau)} - \varphi'(0) > 0, \tag{69}$$

and V *is the steady-state workload.*

An immediate result of Proposition 18 is that from (69) we can derive a bound on the asymptotic bias that does not depend on τ.

Corollary 19. *For any* $\tau \geq 0$,

$$b(\hat{\vartheta}; \tau) := \lim_{m \to \infty} b_m(\hat{\vartheta}; \tau) = \frac{\xi \mathrm{E}[e^{-\psi(\xi)V}\mathbf{1}(V \geq \tau)]}{\psi(\xi)\mathrm{P}(V \geq \tau)} \leq \frac{\xi}{\psi(\xi)}. \tag{70}$$

As $\varphi(\alpha)$ is a convex and increasing function, the term $\xi/\psi(\xi)$ is an increasing function of the sampling rate ξ. In line with earlier observations, it shows that sampling slowly is to be preferred when there is no, or little, cost associated with the sampling duration.

Remark 7. A rough approximation for the expected sample size $\mathrm{E}M(m, \tau)$ required to collect m observations above a threshold τ can be obtained by considering the corresponding stationary tail probability. If the input process is light tailed, then (see [11], Ch. IV),

$$\mathrm{P}(V > \tau) \approx e^{-\omega\tau}, \tag{71}$$

where $\mathrm{E}e^{\omega X(1)} = 1$. If M stationary observations are made then the expected number above τ is $M \cdot \mathrm{P}(V > \tau)$, leading to the approximation $\mathrm{E}M(m, \tau) \cdot \mathrm{P}(V > \tau) \approx m$. We thus obtain

$$\mathrm{E}M(m, \tau) \approx me^{\omega\tau}. \tag{72}$$

6. Simulation analysis

In order to numerically test the performance of the estimation procedure, we carried out simulations of a queue with Lévy input that consists of three processes: $X(t) = C(t) + W(t) + U(t)$, where

- $C(t)$ is Compound Poisson with rate λ and Gamma jumps-size distribution, $\Gamma(\eta, \mu)$,
- $W(t)$ is Brownian motion with drift d and variance σ^2,
- $U(t)$ is a Gamma process (subordinator) with parameters (β, γ).

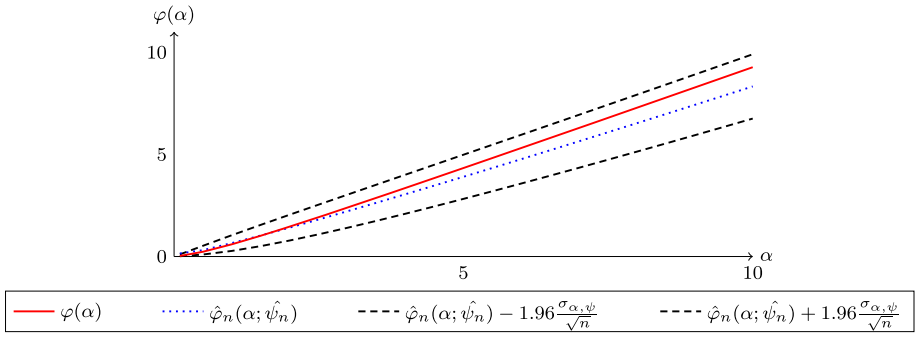

Figure 2. M/M/1 with $\lambda = 0.8$ and $\mu = 1$. The exponent function $\varphi(\alpha)$ (solid red line) and its Z-estimator $\hat{\varphi}_n(\alpha;\hat{\psi}_n)$ (dotted blue line) from a sample of $n = 30$ observations with sampling rate $\xi = 1$. The intermediate estimator $\hat{\psi}_n$ is the MLE of Section 4. Also plotted are the upper and lower bounds of the confidence interval using the normal approximation and asymptotic variance.

The exponent function is then

$$\varphi(\alpha) = \lambda\left[\left(\frac{\mu}{\mu+\alpha}\right)^{\eta} - 1\right] - \alpha d + \frac{1}{2}\alpha^2\sigma^2 + \beta\log\left(\frac{\gamma}{\gamma+\alpha}\right), \tag{73}$$

and the expected input per unit of time is $\mathrm{E}X(1) = -\varphi'(0) = \lambda\eta/\mu + d + \beta/\gamma$. We first examine the performance of the Z-estimator for φ that uses the MLE estimator for $\psi(\xi)$ in the M/M/1 case $(\lambda, \eta, \mu, d, \sigma^2, \beta, \gamma) = (0.8, 1, 1, -1, 0, 0, 0)$. In Figure 2, the real and estimated exponent functions are compared along with the confidence interval based on the asymptotic normal approximation. Even for a moderate sample size of $n = 30$ the estimated function is quite close to the real one. Observe, however, that the estimate is less accurate as α increases. In Figure 3, we illustrate the increase in accuracy when the sample size n increases. Theorem 1 established that the estimator is consistent and here we can see that the bias diminishes very quickly. Increasing the sample size is more important when one wishes to estimate $\varphi(\alpha)$ for high values of α.

In Figure 4, the asymptotic variance $\sigma^2_{\alpha,\xi}$ is plotted as a function of α for different sampling rates. These functions are all increasing and appear to be convex and unbounded. The asymptotic correlation,

$$r(\alpha, 1) := \frac{\sigma^2_{\alpha,1}}{\sqrt{\sigma^2_{\alpha,\xi}\sigma^2_{1,\xi}}} \tag{74}$$

(see Theorem 7), between the estimated function $\hat{\varphi}_n(\alpha;\hat{\psi}_n)$ and $\hat{\varphi}_n(1;\hat{\psi}_n)$ is illustrated in Figure 5. The correlation is a decreasing function of the distance from one, but interestingly does not disappear to zero but to a lower bound (about 0.2 in this example). This should come as no surprise after some thought, as the function is estimated using the same sample of observations for all $\alpha > 0$.

We used the two-step estimation method of $\psi(\xi)$ introduced in Section 5 to test the performance of the estimator $\hat{\varphi}_n$ for Lévy input. The data was simulated for an input process with pa-

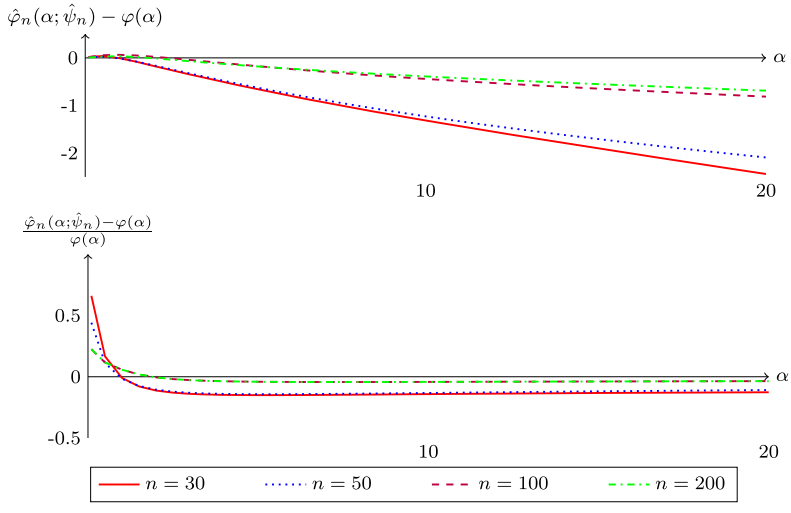

Figure 3. M/M/1 with $\lambda = 0.8$ and $\mu = 1$. The absolute and relative estimation errors of the exponent function, $(\hat{\varphi}_n(\alpha; \hat{\psi}_n) - \varphi(\alpha))/\varphi(\alpha)$, based on sample sizes $n \in \{30, 50, 100, 200\}$ with sampling rate $\xi = 1$. The intermediate estimator $\hat{\psi}_n$ is the MLE of Section 4. Common random numbers were used to simulate the full sample of $n = 200$ and the other estimators use the first n observations.

rameter vector $(\lambda, \eta, \mu, d, \sigma, \beta, \gamma) = (0.2, 1.2, 0.5, -1, 0.1, 1, 5)$ and initial workload $V(0) = 0$. In Figure 6, the estimator is based on $m = 200$ observations with a threshold $\tau = 2$ for different sampling rates. As before, accuracy improves with a slow sampling rate and decreases as α grows. For a fixed sampling rate of $\xi = 1$, Figure 7 illustrates the estimated functions for different threshold levels for the biased intermediate estimation step. Both low and high thresholds give less accurate estimators than the intermediate level, $\tau = 2$ in this example. This is because a low threshold has a large bias, whereas a high threshold has a small bias but a smaller sample size and thus a larger variance.

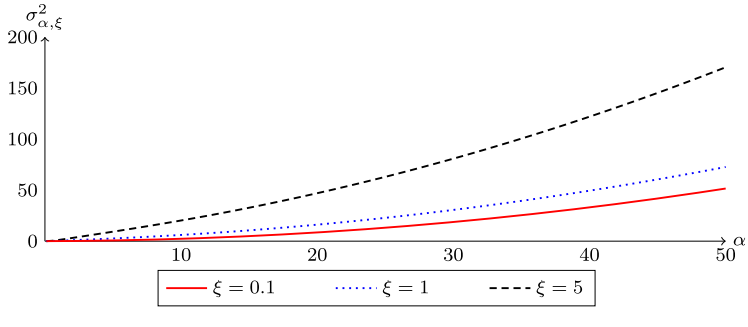

Figure 4. M/M/1 with $\lambda = 0.8$ and $\mu = 1$. The asymptotic variance $\sigma^2_{\alpha,\xi}$ as a function of α for different sampling rates.

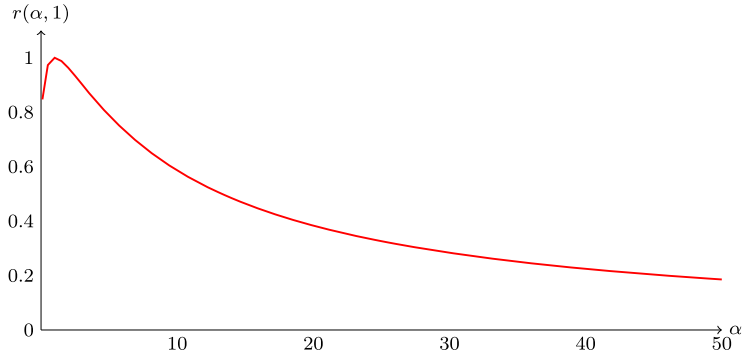

Figure 5. M/M/1 with $\lambda = 0.8$ and $\mu = 1$. The asymptotic correlation $r(\alpha, 1)$, between the estimators $\hat{\varphi}_n(\alpha; \hat{\psi}_n)$ and $\hat{\varphi}_n(1; \hat{\psi}_n)$. The sampling rate is $\xi = 1$.

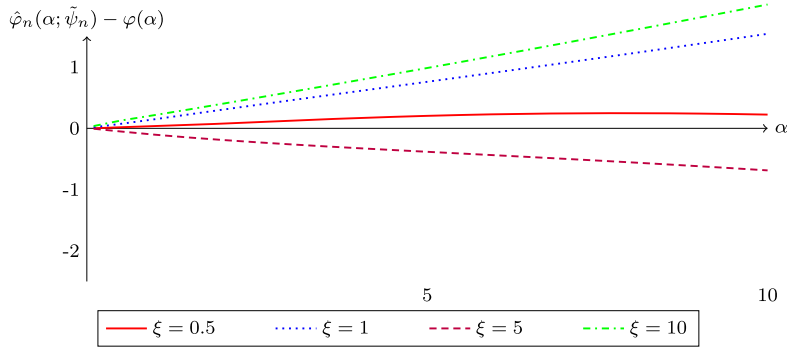

Figure 6. Lévy input with $(\lambda, \eta, \mu, d, \sigma, \beta, \gamma) = (0.2, 1.2, 0.5, -1, 0.1, 1, 5)$. The estimation error of the exponent function, $\hat{\varphi}_n(\alpha; \tilde{\psi}_n) - \varphi(\alpha)$, based on a sample of $m = 200$ observations for different sampling rates $\xi \in \{0.5, 1, 5, 10\}$ and threshold $\tau = 2$.

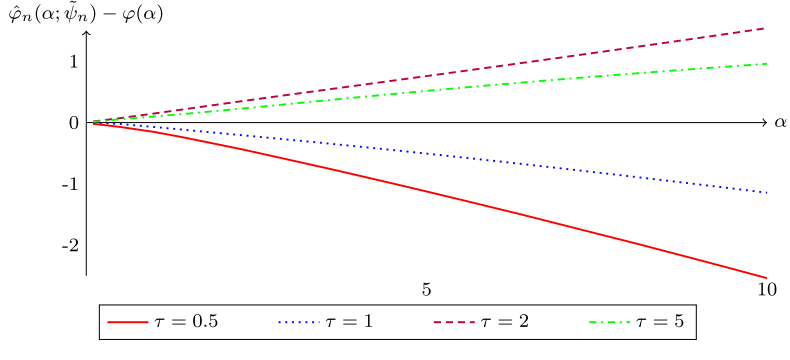

Figure 7. Lévy input with $(\lambda, \eta, \mu, d, \sigma, \beta, \gamma) = (0.2, 1.2, 0.5, -1, 0.1, 1, 5)$. The estimation error of the exponent function, $\hat{\varphi}_n(\alpha; \tilde{\psi}_n) - \varphi(\alpha)$, based on a sample of $m = 200$ observations for different thresholds $\tau \in \{0.5, 1, 2, 5\}$ and sampling rate $\xi = 1$.

7. Discussion and concluding remarks

In this paper, we have studied the problem of statistical inference on the input process to a Lévy-driven queue, where the workload is observed at Poisson times. An estimation method for the characteristic exponent function of the input process has been introduced. The setting is semi-parametric because although the goal is to estimate a function, the methods employed are parametric (namely, method-of-moments and ML-based). Under specific verifiable conditions, consistency and asymptotic normality are proven.

Relation with previous work. In the Introduction we already reflected on the papers [8,20]; we now focus on the relation with the generalized method-of-moments (GMM) framework of [12]. Recall that our main estimation device is the Z-estimator $\hat{\varphi}_n(\alpha)$ which is derived by solving an estimation equation that equates the empirical LST of the workload to the conditional expected value after an exponential sampling time. In this respect, our work is related to [12] that studies GMM estimation of parameters of Markov processes based on data sampled according to a random process which may be time and state dependent. In our setting, a GMM procedure is used to estimate the vector $(\varphi(\alpha_1), \ldots, \varphi(\alpha_p))$ by solving the respective estimation equations (9) for $i = 1, \ldots, p$.

Our estimation equations can be stated in terms of the generator functions used in [12] with a homogeneous Poisson sampling process. However, there is a need to externally estimate the term $\psi(\xi)$; importantly, this cannot be achieved simultaneously using the same estimation equations. This is more than just a technical issue associated with the estimation method, and is inherent to the non-parametric setting of estimating the function φ, as opposed to finite-dimensional parameter estimation. Another issue is that we make inference on the input process without observing it directly but rather from observations of the reflected process. Finally, the asymptotic analysis of [12] only deals with asymptotic normality for the case of a one-dimensional parameter space, where we have derived the multi-dimensional asymptotic distribution of the estimation errors and provided explicit terms for all associated asymptotic (co-)variance terms.

Identifiability for CP input. For the CP case one can in principle identify the λ and $G^*(\alpha)$ separately from the estimates $\hat{\varphi}_n(\alpha; \hat{\psi}_n)$. First one sets, for some large value of α_+,

$$\hat{\lambda}_n = \alpha_+ - \hat{\varphi}_n(\alpha_+; \hat{\psi}_n), \tag{75}$$

and then, for any α,

$$\hat{G}_n^*(\alpha) = 1 + \frac{1}{\hat{\lambda}_n}\big(\hat{\varphi}_n(\alpha; \hat{\psi}_n) - \alpha\big). \tag{76}$$

Parametric special cases. Our estimation method is directly applicable to various parametric special cases using the transient moment equations. For example, when the input is Brownian motion one has $\varphi(\alpha) = -d\alpha + \frac{1}{2}\alpha^2\sigma^2$. In this case, the first two conditional moment equations, (6) and (7), can be used to construct estimators for d and σ^2. For the M/M/1 queue jump sizes follow an exponential distribution with rate μ. Therefore, $G^*(\alpha) = \mu/(\mu + \alpha)$, which yields

$$\varphi(\alpha) = \frac{\alpha(\mu + \alpha - \lambda)}{\mu + \alpha}, \qquad \psi(\xi) = \frac{1}{2}\big(\xi + \lambda - \mu + \sqrt{(\xi + \lambda - \mu)^2 + 4\xi\mu}\big). \tag{77}$$

In Section 4 for the subordinator input case, an MLE for $\psi(\xi)$ was derived. This can be used to estimate the two parameters of an M/M/1, or any other two-parameter formulation for that matter, by using only the first moment equation (6). This can be done similarly for other types of parametric subordinator input such as a Gamma process.

Bootstrap estimation. The asymptotic variance of the estimation errors (in Theorems 3, 7, 12, and 15, as well as Lemmas 5 and 13) are all expressed in terms of the stationary distribution of workload V, which is unknown. For example, the variance of the estimation error of the MLE $\hat{\psi}_n$ in Theorem 12 is

$$I_\xi := \mathrm{E}\left[\frac{\frac{\xi}{\psi(\xi)}e^{-\psi(\xi)V}(\frac{1}{\psi(\xi)}+V)^2}{1-\frac{\xi}{\psi(\xi)}e^{-\psi(\xi)V}}\right], \tag{78}$$

which cannot be directly computed without knowing the distribution of V. A reasonable approach to overcome this issue is to apply a bootstrap procedure: given a sample of size n randomly select a sample of size $m < n$ with replacement and compute the sample mean of the desired term, e.g. I_ξ, and repeat the above a large number of times and compute the mean of all results. Such a bootstrap procedure is known to be consistent. As a consequence, for large n and m the method has good accuracy. We refer to e.g. [35], Ch. 23, for general results, and [19] for a specific implementation dealing with the stationary workload of an M/G/1 queue.

Future research. There are several interesting questions that are left open in this work. First of all, the question whether a consistent estimator for $\psi(\xi)$ can be constructed for the general case. Although simulation analysis suggests that in practice the biased estimator performs well, it is still of interest to find a consistent estimator for theoretical and practical efficiency purposes. Another issue is that the object of estimation is the function φ but all asymptotic results are in terms of the function at a finite collection of points. A natural extension would be to derive functional limit theorems, for example, uniform-consistency and a FCLT approximation for the estimation error as a Gaussian process. In Section 6, we observed that the asymptotic variance appears to be increasing and unbounded with α which may lead to technical difficulties, and in particular it is likely that the estimator is not uniformly-consistent.

Applications. We conclude this paper by mentioning a few applications for which the model and methods presented here may be useful.

- *Optimal sampling rate*. As we saw, explicitly in Section 4 and numerically in Section 6, the accuracy of the estimators improves as the sampling rate decreases. This, however, comes at a cost of a longer sampling duration. In applications where there is a cost associated with the duration of the learning period of the system, this leads to the problem of finding an optimal sampling rate that balances this tradeoff. Our results can be used to derive an approximate solution.
- *Stability detection*. Throughout this work, we assumed that $\mathrm{E}X(1) < 0$ and consequently the workload process is stable. However when the input is unknown, we do not know *a priori* whether $\mathrm{E}X(1) < 0$. For example, if the processing speed can be controlled in an M/G/1 queue but the input is unknown then it is possible that for some processing speeds the queue is unstable. In this case, a likelihood ratio test can be constructed using the setup of Section 4. In the general case, the biased estimator $\hat{\vartheta}_n$ for $-\mathrm{E}X(1)$ (Section 5) can be used to set up a threshold rule for detecting instability.

- *Dynamic pricing*. Suppose that the server can charge an admission price with the goal of maximizing social welfare or revenue. The optimal pricing scheme obviously depends on the input distribution [22], which may not be known. This naturally leads to a model with an exploration-exploitation element, and using our framework the exploration step can be carried out for transient observations without relying on any approximations. The extension to the dynamic setting requires a state-dependent input model such as [6].
- *Testing distributional assumptions*. An issue that is related to the two previous points is that often controllable system parameters are chosen with the goal of maximizing specific performance measures. This is typically done under some modeling assumptions. Our framework enables testing whether the input to the system is consistent with the modeling assumptions (as long as the input alternatives are within the class of spectrally positive Lévy processes).
- *Service level monitoring*. Service systems are often obliged to provide performance guarantees. For example, when an Internet Service Provider commits to certain bandwidth, it is important for users and regulators to be able to estimate the performance of the system in order to monitor whether the contract is upheld or not [21], Ch. IX. Our method allows to estimate performance measures, such as moments and tail probabilities, using a transient sample of observations.

Acknowledgments

We wish to thank Offer Kella for his valuable advice and comments and two anonymous reviewers for their helpful comments. This research was carried out as part of the NWO Gravitation project NETWORKS, grant number 024.002.003.

Supplementary Material

Online appendix. Estimating the input of a Lévy-driven queue by Poisson sampling of the workload process (DOI: 10.3150/19-BEJ1109SUPP; .pdf). The appendix provides the proofs for several lemmas and theorems.

References

[1] Afèche, P. and Ata, B. (2013). Bayesian dynamic pricing in queueing systems with unknown delay cost characteristics. *Manufacturing & Service Operations Management* **15** 292–304.

[2] Antunes, N., Jacinto, G. and Pacheco, A. (2014). Probing a M/G/1 queue with general input and service times. *ACM SIGMETRICS Perform. Eval. Rev.* **41** 34–36.

[3] Antunes, N., Jacinto, G., Pacheco, A. and Wichelhaus, C. (2016). Estimation of the traffic intensity in a piecewise-stationary $M_t/G_t/1$ queue with probing. *ACM SIGMETRICS Perform. Eval. Rev.* **44** 3–5.

[4] Armony, M. and Plambeck, E.L. (2005). The impact of duplicate orders on demand estimation and capacity investment. *Manage. Sci.* **51** 1505–1518.

[5] Asanjarani, A., Nazarathy, Y. and Pollett, P.K. (2017). Parameter and state estimation in queues and related stochastic models: A bibliography. Available at arXiv:1701.08338.

[6] Bekker, R., Boxma, O.J. and Resing, J.A.C. (2009). Lévy processes with adaptable exponent. *Adv. in Appl. Probab.* **41** 177–205. MR2514950

[7] Belomestny, D. (2011). Statistical inference for time-changed Lévy processes via composite characteristic function estimation. *Ann. Statist.* **39** 2205–2242. MR2893866

[8] Bingham, N.H. and Pitts, S.M. (1999). Nonparametric inference from $M/G/1$ busy periods. *Comm. Statist. Stochastic Models* **15** 247–272. MR1683605

[9] Blanghaps, N., Nov, Y. and Weiss, G. (2013). Sojourn time estimation in an M/G/∞ queue with partial information. *J. Appl. Probab.* **50** 1044–1056. MR3161372

[10] Buchmann, B. and Grübel, R. (2003). Decompounding: An estimation problem for Poisson random sums. *Ann. Statist.* **31** 1054–1074. MR2001642

[11] Dębicki, K. and Mandjes, M. (2015). *Queues and Lévy Fluctuation Theory*. *Universitext*. Springer. MR3379923

[12] Duffie, D. and Glynn, P. (2004). Estimation of continuous-time Markov processes sampled at random time intervals. *Econometrica* **72** 1773–1808. MR2095532

[13] Ferguson, T.S. (1996). *A Course in Large Sample Theory*. *Texts in Statistical Science Series*. London: CRC Press. MR1699953

[14] Glynn, P.W., Melamed, B. and Whitt, W. (1993). Estimating customer and time averages. *Oper. Res.* **41** 400–408.

[15] Goldenshluger, A. (2016). Nonparametric estimation of the service time distribution in the M/G/∞ queue. *Adv. in Appl. Probab.* **48** 1117–1138. MR3595768

[16] Goldenshluger, A. (2018). The $M/G/\infty$ estimation problem revisited. *Bernoulli* **24** 2531–2568. MR3779694

[17] Gugushvili, S. (2009). Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametr. Stat.* **21** 321–343. MR2530929

[18] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application*. New York-London: Academic Press [Harcourt Brace Jovanovich, Publishers]. Probability and Mathematical Statistics. MR0624435

[19] Hansen, M.B. (2005). Nonparametric estimation of the stationary M/G/1 workload distribution function. In *Proceedings of the* 37*th Conference on Winter Simulation* 869–877. Winter Simulation Conference.

[20] Hansen, M.B. and Pitts, S.M. (2006). Nonparametric inference from the $M/G/1$ workload. *Bernoulli* **12** 737–759. MR2248235

[21] Hassin, R. (2016). *Rational Queueing*. Boca Raton, FL: CRC Press. MR3467608

[22] Hassin, R. and Haviv, M. (2003). *To Queue or Not to Queue*: *Equilibrium Behavior in Queueing Systems*. *International Series in Operations Research & Management Science* **59**. Boston, MA: Kluwer Academic. MR2006433

[23] Heijmans, R.D.H. and Magnus, J.R. (1986). Consistent maximum-likelihood estimation with dependent observations. The general (nonnormal) case and the normal case. *J. Econometrics* **32** 253–285. MR0864928

[24] Heijmans, R.D.H. and Magnus, J.R. (1986). On the first-order efficiency and asymptotic normality of maximum likelihood estimators obtained from dependent observations. *Stat. Neerl.* **40** 169–188. MR0868075

[25] Heyde, C.C. (1997). *Quasi-Likelihood and Its Application*: *A General Approach to Optimal Parameter Estimation*. *Springer Series in Statistics*. New York: Springer. MR1461808

[26] Kella, O., Boxma, O. and Mandjes, M. (2006). A Lévy process reflected at a Poisson age process. *J. Appl. Probab.* **43** 221–230. MR2225062

[27] Kyprianou, A.E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. *Universitext*. Berlin: Springer. MR2250061

[28] Neumann, M.H. and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15** 223–248. MR2546805

[29] Novak, A., Taylor, P. and Veitch, D. (2006). The distribution of the number of arrivals in a subinterval of a busy period of a single server queue. *Queueing Syst*. **53** 105–114. MR2243923

[30] Pitts, S.M. (1994). Nonparametric estimation of the stationary waiting time distribution function for the $GI/G/1$ queue. *Ann. Statist*. **22** 1428–1446. MR1311982

[31] Ravner, L., Boxma, O. and Mandjes, M. (2019). Supplement to "Estimating the input of a Lévy-driven queue by Poisson sampling of the workload process." DOI:10.3150/19-BEJ1109SUPP.

[32] Ross, J.V., Taimre, T. and Pollett, P.K. (2007). Estimation for queues from queue length data. *Queueing Syst*. **55** 131–138. MR2311210

[33] Schweer, S. and Wichelhaus, C. (2015). Nonparametric estimation of the service time distribution in the discrete-time $GI/G/\infty$ queue with partial information. *Stochastic Process. Appl*. **125** 233–253. MR3274698

[34] Smith, D.E. and Li, M. (2002). Estimating data delays from Poisson probe delays, Aug. 6 2002. US Patent 6,430,160.

[35] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247

[36] Yao, D.D. (1992). On Wolff's PASTA martingale. *Oper. Res*. **40** S352–S355.

Received August 2018 and revised January 2019