MARKOVIAN MULTIVARIATE HAWKES POPULATION PROCESSES: EFFICIENT EVALUATION OF MOMENTS

RAVIAR S. KARIM, ROGER J. A. LAEVEN, AND MICHEL MANDJES

ABSTRACT. We provide probabilistic and computational results on Markovian multivariate Hawkes processes and induced population processes. By applying the Markov property, we characterize in closed form a joint transform, bijective to the probability distribution, of the population process and its underlying intensity process. We demonstrate a method that exploits the transform to obtain analytic expressions for transient and stationary multivariate moments of any order, as well as auto- and cross-covariances. We reveal a nested sequence of block matrices that yields the moments in explicit form and brings important computational advantages. We also establish the asymptotic behavior of the intensity of the multivariate Hawkes process in its nearly unstable regime, under a specific parameterization. In extensive numerical experiments, we analyze the computational complexity, accuracy and efficiency of the established results.

KEYWORDS. Hawkes processes ∘ mutual excitation ∘ Markov processes ∘ transform analysis ∘ moment computations

ACKNOWLEDGEMENTS AND AFFILIATIONS. We are very grateful to Jan Magnus for helpful comments. RK and RL are with the Dept. of Quantitative Economics, University of Amsterdam. RL is also with EURANDOM, Eindhoven University of Technology, and with CENTER, Tilburg University; his research was funded in part by the Netherlands Organization for Scientific Research under grants NWO VIDI 2009 and NWO VICI 2019/20. MM is with the Korteweg-de Vries Institute for Mathematics, University of Amsterdam. MM is also with EURANDOM, Eindhoven University of Technology, and with the Amsterdam Business School, University of Amsterdam; his research was funded in part the Netherlands Organization for Scientific Research under the Gravitation project NETWORKS, grant 024.002.003.

Email: `raviar3@hotmail.com`, `r.j.a.laeven@uva.nl`, `m.r.h.mandjes@uva.nl`.

This document can be found at: `arXiv:2506.08775`

Contents

1. Introduction

Since the onset of globalization, the mechanisms according to which events spread have become increasingly complex. Events involving disease contagion, financial panic, news that goes viral, are all subject to forms of propagation that occur through time and space, be it across human populations, equity markets, or media outlets. A multitude of mathematical models have been proposed to describe the corresponding underlying dynamics.

Multivariate point processes constitute one such class of models that describe the random nature of the arrival, and subsequent spread, of events, in the time as well as space domain. In particular, the subclass of multivariate Hawkes processes ([14, 15]) provides a rich structure that is capable of capturing contagious dynamics. These Hawkes processes allow for flexible dependencies of events, due to the inherent feedback mechanisms known as *self-* and *cross-excitation*.

Recently, Hawkes processes have been increasingly used as the arrival process of infinite-server queues ([12, 18]). Such infinite-server queues can be viewed as population processes, where individuals are born (i.e., arrive) according to a Hawkes process and die (i.e., leave) after a random time. A related, relevant motivation to study a Hawkes-fed infinite-server queue is to account for infected and recovered subjects in a population. Hawkes processes have been widely applied in epidemiology, in conjunction with SIR models ([23]) and dynamic contagion models, e.g., in the context of the COVID-19 pandemic ([5]). Throughout this paper, we use the terminology of population processes and infinite-server queues interchangeably.

An important strand of research considers a *multivariate* framework to allow for cross-exciting dynamics between subpopulations (for instance, residing at different geographic locations). Also in various other domains this type of model offers a natural framework, for example when considering the number of simultaneous online visitors of a specific website or to describe social interaction. In the operations management literature, (multivariate) Hawkes processes have gained significant interest over the past decade, e.g. in the context of modeling customer support contact centers ([4, 10]), in the context of jump propagation in spot and options markets ([13, 2]), and in relation to online advertising ([24]).

When analyzing infinite-server queues with multivariate Hawkes input, the main challenge lies in unraveling, and tractably representing, its inherently complex probabilistic structure. In this paper our main objective is to devise computational techniques to efficiently and accurately evaluate moments. We consider the Markovian case, i.e., the case in which the excitation functions are exponentially decaying; see ([14, 21]). The Markov property is used to obtain a characterization of the transform of the joint process, that is, the Hawkes intensity process and the infinite-server queue, in terms of a system of differential equations.

Contributions. This paper makes five contributions. We start by deriving in closed form a joint transform — bijective to the joint probability distribution — of the multivariate population process and its underlying intensity process, exploiting their Markovian nature.

We allow for general distributions of the intensity jump sizes and allow the processes to be evaluated at potentially multiple future time instances, thus characterizing all cross-sectional and temporal probabilistic features. Second, we employ this joint transform to obtain explicit, recursive expressions for both the transient and stationary multivariate cross-moments. Third, we show that the higher-order transient and stationary moments can be obtained in closed form from a nested sequence of block matrices, having important computational advantages. Fourth, for a specifically chosen set of parameters, we establish the limiting behavior of the underlying intensity process of the multivariate population process in the practically relevant, nearly unstable situation, where the stability condition is close to being violated; cf. the *heavy-traffic regime* in queueing theory. We thus substantially extend earlier results pertaining to the case that the underlying Hawkes process is single-dimensional ([9, 12, 18, 11]); cf. also [6].

The final contribution concerns the use of our analytic findings when devising efficient computational techniques for evaluating moments. To this end, we analyze in numerical experiments the methods developed in this paper and assess their accuracy and efficiency. The methods we develop show superior performance in terms of speed and accuracy when compared to the computational alternative of applying finite difference schemes to the joint transform. Our computational approach is fast and accurate: it provides near-instant response that is exact up to machine precision. When using the nested block matrices, it has the attractive feature that the computation speed does not depend on the value of the considered time horizon. Moreover, when compared to the simulation alternative, a substantial number of Monte Carlo simulation runs is needed to sufficiently precisely approximate the object of interest, so that the performance gain is even orders of magnitude larger. The computer codes that implement the methods developed in this paper are available at `https://github.com/RaviarKarim/HawkesMarkov`.

In the computation of the transient and stationary moments, we focus on two settings, which are in a sense each others dual, namely the bivariate setting with moments of arbitrary order and the multivariate model with moments up to second order. The generic recursive structure for computing the transient moments resembles the one by which their stationary counterparts can be evaluated, with one crucial difference: in the transient setting we obtain recursive systems of non-homogeneous linear differential equations, whereas in the stationary setting we obtain recursive systems of linear algebraic equations.

An important application that our results make possible, is moment-based estimation of multivariate Hawkes processes. Such an estimation approach requires evaluating a collection of moments, auto-covariances and cross-covariances a (very) large number of times, for the parameter vector proposed at each new iteration of the optimization routine. Existing approaches to evaluating moments and other distributional characteristics are computationally prohibitively expensive for this purpose. One might employ the explicit approximate moments, obtained from the infinitesimal Markov generator using operator methods and Taylor expansions applied to short time intervals, as in [1]. This approach, however, requires data sampled at least at the daily frequency. Our approach is based on exact expressions of the moments, auto-covariances and cross-covariances over time intervals of arbitrary length, and their numerical evaluation remains fast and accurate. Another advantage of our approach is that it facilitates comparative statics, i.e., it provides an efficient, tractable link between the objects of interest and the parameters of the process.

Organization. This paper is organized as follows. In Section 2, we introduce the Markovian multivariate Hawkes process and the induced population process. In Section 3, we derive a characterization of the joint transform using the Markov property of the process. Subsequently, we use this result to obtain relations for the transient and stationary moments. In Section 4, we show the underlying recursive structure in the computation of moments of certain order and dimension. In Section 5, we focus on the bivariate setting, revealing a nested block structure of matrices that characterize the moments, enabling fast computation. In Section 6, we analyze the nearly unstable case for the intensity of the process. Section 7 provides our numerical experiments. Conclusions are in Section 8. Proofs, technical derivations and details, and additional numerical experiments are relegated to six Appendices.

2. Multivariate Hawkes Populations

In this section we define the multivariate Hawkes process, by means of the associated conditional intensity function, as well as its induced population process. Hawkes processes, first introduced in a series of papers [14, 15, 16], are a class of point processes that exhibit self-exciting behavior, in the sense that the current value of the associated intensity function depends on the history of the point process. In this paper, we focus on multivariate Hawkes processes of the Markovian type.

We consider a d-dimensional point process $\boldsymbol{N}(\cdot) \equiv (\boldsymbol{N}(t))_{t\in\mathbb{R}_+} = ((N_1(t), \ldots, N_d(t)))^\top_{t\in\mathbb{R}_+}$, which records the number of points in each component $N_i(t)$, with $i \in [d] := \{1, \ldots, d\}$, up to and including time t. It is well known that a point process is characterized by its conditional intensity function $\boldsymbol{\lambda}(t) = (\lambda_1(t), \ldots, \lambda_d(t))^\top$, see [8, Chapter 7].

Definition 1. *A Markovian multivariate Hawkes process is a point process $\boldsymbol{N}(\cdot)$, with $\boldsymbol{N}(0) = \mathbf{0}$, whose components $N_i(\cdot)$ satisfy*

$$\begin{aligned}
&\mathbb{P}(N_i(t+\Delta) - N_i(t) = 0 \,|\, \mathcal{F}_t) = 1 - \lambda_i(t)\Delta + o(\Delta),\\
&\mathbb{P}(N_i(t+\Delta) - N_i(t) = 1 \,|\, \mathcal{F}_t) = \lambda_i(t)\Delta + o(\Delta),\\
&\mathbb{P}(N_i(t+\Delta) - N_i(t) > 1 \,|\, \mathcal{F}_t) = o(\Delta),
\end{aligned} \tag{1}$$

as $\Delta \downarrow 0$, with $\mathcal{F}_t = \sigma(\boldsymbol{N}(s) \,:\, s \leqslant t)$ denoting the natural filtration generated by $\boldsymbol{N}(\cdot)$. Here, each component $\lambda_i(t)$ of the intensity function satisfies the mean-reverting dynamics

$$\mathrm{d}\lambda_i(t) = \alpha_i(\overline{\lambda}_i - \lambda_i(t))\,\mathrm{d}t + \sum_{j=1}^{d} B_{ij}(t)\,\mathrm{d}N_j(t), \tag{2}$$

where $\lambda_i(0) = \overline{\lambda}_i \geqslant 0$, $\alpha_i \geqslant 0$, and for each $i, j \in [d]$, $(B_{ij}(t))_t$ is a sequence of independent random variables, distributed as the generic non-negative random variable B_{ij}.

The intuitive explanation behind Definition 1 goes as follows. The constants $\overline{\lambda}_i$ are referred to as the *base rates*. When a point is generated in component $j \in [d]$, $N_j(t)$ increases by one and makes intensity $\lambda_i(t)$, for all $i \in [d]$, jump by a value $B_{ij}(t)$ that is distributed as the random variable B_{ij}. This jump in the intensity $\lambda_i(t)$ caused by $B_{ij}(t)$ increases the probability of new points being generated in component i, thereby more likely to increase $N_i(t)$. After the jump has occurred, the intensity $\lambda_i(t)$ decays exponentially with rate α_i, in the direction of the base rate $\overline{\lambda}_i$. These jumps in intensities and the subsequent decay is what makes points cluster across time and space. That is, for any $j \in [d]$, when points in $N_j(t)$ cause $\lambda_j(t)$ to jump, there is a pure temporal effect and we speak of *self-excitation*, while an effect on $\lambda_i(t)$, with $i \neq j$, has an additional spatial effect and we speak of *cross-excitation*.

It is well-known that with exponential decay the joint process $(\boldsymbol{N}(\cdot), \boldsymbol{\lambda}(\cdot))$ is a Markov process, see e.g. [19, 20, 21]. Furthermore, by Itô's Lemma applied to $f(t, \lambda_i(t)) = e^{\alpha_i t}\lambda_i(t)$, Eqn. (2) can alternatively be expressed as

$$\lambda_i(t) = \overline{\lambda}_i + \sum_{j=1}^{d} \int_0^t B_{ij}(s)\, e^{-\alpha_i(t-s)}\, \mathrm{d}N_j(s), \tag{3}$$

where we used that $\boldsymbol{N}(0) = \boldsymbol{0}$, implying $\lambda_i(0) = \overline{\lambda}_i$. The exponential function $g_i(t) := e^{-\alpha_i t}$ in Eqn. (3) is known in the literature as the *decay function*. We emphasize that only exponentially decaying $g_i(\cdot)$ render the joint process $(\boldsymbol{N}(\cdot), \boldsymbol{\lambda}(\cdot))$ a Markov process.

To ensure stability of the multivariate Hawkes process, [15, 3] show that we must impose a stability condition. In what follows, we assume that this stability condition applies.

Assumption 1. *With $\rho(\cdot)$ denoting the spectral radius of a matrix, assume that $\rho(\boldsymbol{H}) < 1$, where the matrix $\boldsymbol{H} = (h_{ij})_{i,j\in[d]}$ has elements*

$$h_{ij} \equiv \mathbb{E}[B_{ij}] \int_0^\infty e^{-\alpha_i t}\, \mathrm{d}t = \mathbb{E}[B_{ij}]/\alpha_i. \tag{4}$$

In the model discussed in this paper, the Markovian multivariate Hawkes process will serve as the arrival process of a multivariate population process. Our overarching goal is to compute various quantities pertaining to the (joint) distribution of the arrival process and the population process. A similar setup has been considered in [12, 18], where the univariate Hawkes process serves as the arrival process for an infinite-server queue in which arriving customers reside for independent and identically distributed (i.i.d.) amounts of time (often assumed exponentially distributed). The following definition generalizes that framework to the multivariate setting.

Definition 2. *Let $\boldsymbol{N}(\cdot)$ be a Markovian multivariate Hawkes process as given in Definition 1. Define the associated Hawkes population process $\boldsymbol{Q}(\cdot)$, with $\boldsymbol{Q}(0) = \boldsymbol{0}$, by setting, for any component i and $t \geqslant 0$,*

$$Q_i(t) := \int_0^t \mathbf{1}_{\{E_i(s) > t-s\}}\, \mathrm{d}N_i(s), \tag{5}$$

where $(E_i(s))_s$ is a sequence of independent random variables, exponentially distributed with parameter μ_i, also independent of the multivariate arrival process $\boldsymbol{N}(t)$.

The above definition entails that each arrival in component $i \in [d]$ remains in the system for an exponentially distributed amount of time. In demography, one can think of $Q_i(\cdot)$ as the number of individuals in a subpopulation $i \in [d]$, where each $E_i(s)$ models the lifetime of an individual in this subpopulation. In epidemiology, $Q_i(\cdot)$ would represent the number of individuals infected by a disease at location $i \in [d]$, where $E_i(s)$ would model the duration from infection to recovery (or death). In queuing terminology, this type of system can be interpreted as a special case of an infinite-server queue. We note that if $\mu_i \equiv 0$, then no points ever depart from component i, and hence $Q_i(\cdot) \equiv N_i(\cdot)$. However, in some of the expressions we will encounter in this paper one must take proper care of taking the limit $\mu_i \downarrow 0$, so as to avoid dividing by zero; note that if $\mu_i = 0$, then $Q_i(\cdot)$ eventually grows unbounded.

The application of our methodology to the class of multivariate Hawkes population processes, as performed in the present paper, may be viewed as a "proof of principle". We would like to stress that our methodology can, in principle, be applied to any general multivariate Markov process.

3. Transform and Joint Moments

The objective of this section is to analyze the joint transform of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$ for any fixed $t \in \mathbb{R}_+$, and to use this transform to obtain an algorithm to determine the corresponding (joint) moments. To this end, define, for given initial values $\boldsymbol{Q}(t_0) = \boldsymbol{Q}_0 \in \mathbb{N}_0^d$ and $\boldsymbol{\lambda}(t_0) = \boldsymbol{\lambda}_0 \in \mathbb{R}_+^d$ for some $0 \leqslant t_0 < t$, the conditional joint transform

$$\zeta_{t_0}(t, \boldsymbol{s}, \boldsymbol{z}) = \mathbb{E}_{t_0}\big[\prod_{i=1}^d z_i^{Q_i(t)} e^{-s_i\lambda_i(t)}\big] := \mathbb{E}\big[\prod_{i=1}^d z_i^{Q_i(t)} e^{-s_i\lambda_i(t)} \,|\, \boldsymbol{Q}(t_0) = \boldsymbol{Q}_0, \boldsymbol{\lambda}(t_0) = \boldsymbol{\lambda}_0\big], \tag{6}$$

where $t \geqslant 0$, $\boldsymbol{s} \in \mathbb{R}_+^d$ and $\boldsymbol{z} \in [-1,1]^d$. In the specific case that $t_0 = 0$, with the assumed initial conditions $\boldsymbol{Q}(0) = \boldsymbol{0}$ and $\boldsymbol{\lambda}(0) = \overline{\boldsymbol{\lambda}}$, we define $\zeta(t, \boldsymbol{s}, \boldsymbol{z}) \equiv \zeta_0(t, \boldsymbol{s}, \boldsymbol{z})$ as

$$\zeta(t, \boldsymbol{s}, \boldsymbol{z}) : = \mathbb{E}\big[\prod_{i=1}^d z_i^{Q_i(t)} e^{-s_i\lambda_i(t)}\big] = \mathbb{E}\big[\prod_{i=1}^d z_i^{Q_i(t)} e^{-s_i\lambda_i(t)} \,|\, \boldsymbol{Q}(0) = \boldsymbol{0}, \boldsymbol{\lambda}(0) = \overline{\boldsymbol{\lambda}}\big], \tag{7}$$

where the expectation operator $\mathbb{E}[\cdot]$ is understood as the conditional $\mathbb{E}_0[\cdot]$.

3.1. **Transform characterization.**

The following theorem, proven in Appendix A, identifies the joint transform $\zeta_{t_0}(t, \boldsymbol{s}, \boldsymbol{z})$. Define $\beta_j(\boldsymbol{s}) := \mathbb{E}[e^{-\boldsymbol{s}^\top \boldsymbol{B}_j}] = \mathbb{E}[\exp(-\sum_{i=1}^d s_i B_{ij})]$.

Theorem 1. *Fix $t \in \mathbb{R}_+$, and assume $\boldsymbol{Q}(t_0) = \boldsymbol{Q}_0 \in \mathbb{N}_0^d$ and $\boldsymbol{\lambda}(t_0) = \boldsymbol{\lambda}_0 \in \mathbb{R}_+^d$ for some $0 \leqslant t_0 < t$. Then, for any $\boldsymbol{z} \in [-1,1]^d$, $\boldsymbol{s} \in \mathbb{R}_+^d$,*

$$\zeta_{t_0}(t, \boldsymbol{s}, \boldsymbol{z}) = \prod_{j=1}^d \widehat{z}_j(t_0)^{Q_{j,0}} \exp\Big(- \tilde{s}_j(t)\lambda_{j,0} - \overline{\lambda}_j \alpha_j \int_{t_0}^t \tilde{s}_j(u) \mathrm{d}u\Big), \tag{8}$$

where, for $t_0 \leqslant u \leqslant t$ and $j \in [d]$, the functions $\widehat{z}_j(\cdot)$ and $\tilde{s}_j(\cdot)$ satisfy

$$\begin{aligned} &\widehat{z}_j(u) = 1 + (z_j - 1)e^{-\mu_j(t-u)}, \\ &\frac{\mathrm{d}\tilde{s}_j(u)}{\mathrm{d}u} + \alpha_j \tilde{s}_j(u) + \big(1 + (z_j - 1)e^{-\mu_j(u-t_0)}\big)\beta_j(\tilde{\boldsymbol{s}}(u)) - 1 = 0, \end{aligned} \tag{9}$$

with boundary condition $\tilde{s}_j(t_0) = s_j$.

Corollary 1. *Fix $t \in \mathbb{R}_+$, and let $\boldsymbol{Q}(0) = \boldsymbol{0}$, $\boldsymbol{\lambda}(0) = \overline{\boldsymbol{\lambda}}$. Then, for any $\boldsymbol{z} \in [-1,1]^d$, $\boldsymbol{s} \in \mathbb{R}^d_+$,*

$$\zeta(t, \boldsymbol{s}, \boldsymbol{z}) = \prod_{j=1}^{d} \exp\Big(-\overline{\lambda}_j \tilde{s}_j(t) - \overline{\lambda}_j \alpha_j \int_0^t \tilde{s}_j(v)\mathrm{d}v \Big). \tag{10}$$

Here, for $j \in [d]$ and $v \in [0,t]$, the functions $\tilde{s}_j(\cdot)$ satisfy, with boundary condition $\tilde{s}_j(0) = s_j$,

$$\frac{\mathrm{d}\tilde{s}_j(v)}{\mathrm{d}v} + \alpha_j \tilde{s}_j(v) + (1 + (z_j - 1)e^{-\mu_j v})\beta_j(\tilde{\boldsymbol{s}}(v)) - 1 = 0. \tag{11}$$

Clearly, using the expression presented in Theorem 1, in principle any conditional joint moment can be obtained. Indeed, for any $n_{\lambda_i}, n_{Q_i} \in \mathbb{N}_0$, we have

$$\frac{\mathrm{d}^{n_{\lambda_1}}}{\mathrm{d}s_1^{n_{\lambda_1}}} \cdots \frac{\mathrm{d}^{n_{\lambda_d}}}{\mathrm{d}s_d^{n_{\lambda_d}}} \frac{\mathrm{d}^{n_{Q_1}}}{\mathrm{d}z_1^{n_{Q_1}}} \cdots \frac{\mathrm{d}^{n_{Q_d}}}{\mathrm{d}z_d^{n_{Q_d}}} \zeta_{t_0}(t, \boldsymbol{s}, \boldsymbol{z}) \bigg|_{\substack{\boldsymbol{s}=\boldsymbol{0} \\ \boldsymbol{z}=\boldsymbol{1}}} = \mathbb{E}_{t_0}\Big[\prod_{i=1}^{d} (-1)^{n_{\lambda_i}} \lambda_i(t)^{n_{\lambda_i}} Q_i(t)^{[n_{Q_i}]} \Big], \tag{12}$$

i.e., an object composed from (standard) moments of $\lambda_i(t)$ and *reduced* moments of $Q_i(t)$. Note that in (12) we have used the standard Pochhammer notation: for integers m and n we denote $m^{[n]} := m(m-1)\cdots(m-n+1)$, by convention setting $m^{[0]} := 1$.

In the above, we focused on identifying transforms pertaining to a single point in time. Using similar methods, however, it is possible to derive the transform of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$ and $(\boldsymbol{Q}(t+\tau), \boldsymbol{\lambda}(t+\tau))$ jointly, for some $\tau > 0$. More precisely, the following theorem (proven in Appendix A) identifies, for $\boldsymbol{y}, \boldsymbol{z} \in [-1,1]^d$ and $\boldsymbol{r}, \boldsymbol{s} \in \mathbb{R}^d_+$, the object

$$\zeta_\tau(t, \boldsymbol{r}, \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{z}) := \mathbb{E}\Big[\prod_{i=1}^{d} y_i^{Q_i(t)} e^{-r_i \lambda_i(t)} z_i^{Q_i(t+\tau)} e^{-s_i \lambda_i(t+\tau)} \Big], \tag{13}$$

where as before, $\mathbb{E}[\cdot]$ is understood as $\mathbb{E}_0[\cdot]$, with $\boldsymbol{Q}(0) = \boldsymbol{0}$ and $\boldsymbol{\lambda}(0) = \overline{\boldsymbol{\lambda}}$. In addition, $\boldsymbol{y} \odot \boldsymbol{z}$ is the component-wise product of the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$.

Theorem 2. *Fix $t, \tau \in \mathbb{R}_+$, and let $\boldsymbol{Q}(0) = \boldsymbol{0}$, $\boldsymbol{\lambda}(0) = \overline{\boldsymbol{\lambda}}$. Then, for any $\boldsymbol{y}, \boldsymbol{z} \in [-1,1]^d$, $\boldsymbol{r}, \boldsymbol{s} \in \mathbb{R}^d_+$,*

$$\begin{aligned} \zeta_\tau(t, \boldsymbol{r}, \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{z}) &= \zeta(t, \boldsymbol{y} \odot \widehat{\boldsymbol{z}}(t), \boldsymbol{r} + \tilde{\boldsymbol{s}}(t+\tau)) \prod_{j=1}^{d} \exp\Big(-\overline{\lambda}_j \alpha_j \int_t^{t+\tau} \tilde{s}_j(u)\mathrm{d}u \Big) \\ &= \prod_{j=1}^{d} \exp\Big(-\overline{\lambda}_j \tilde{r}_j(t) - \overline{\lambda}_j \alpha_j \int_0^t \tilde{r}_j(v)\mathrm{d}v - \overline{\lambda}_j \alpha_j \int_t^{t+\tau} \tilde{s}_j(u)\mathrm{d}u \Big). \end{aligned} \tag{14}$$

Here $\zeta(\cdot)$ is given by Eqn. (10)*, and, for $j \in [d]$, the functions $\widehat{z}_j(\cdot)$, $\tilde{s}_j(\cdot)$ and $\tilde{r}_j(\cdot)$ satisfy*

$$\begin{aligned}
&\widehat{z}_j(u) = 1 + (z_j - 1)e^{-\mu_j(t+\tau-u)},\\
&\frac{\mathrm{d}\tilde{s}_j(u)}{\mathrm{d}u} + \alpha_j \tilde{s}_j(u) + \big(1 + (z_j - 1)e^{-\mu_j(u-t)}\big)\beta_j(\tilde{\boldsymbol{s}}(u)) - 1 = 0,\\
&\frac{\mathrm{d}\tilde{r}_j(v)}{\mathrm{d}v} + \alpha_j \tilde{r}_j(v) + \big(1 + (y_j - 1)e^{-\mu_j v} + y_j(z_j - 1)e^{-\mu_j(v+\tau)}\big)\beta(\tilde{\boldsymbol{r}}(v)) - 1 = 0,
\end{aligned} \tag{15}$$

with boundary condition $\tilde{s}_j(t) = s_j$ and $\tilde{r}_j(0) = r_j + \tilde{s}_j(t+\tau)$, and where $0 \leqslant v \leqslant t$ and $t \leqslant u \leqslant t + \tau$.

Observe that the result of Theorem 2 can be extended to include arbitrarily many time points $t < t_1 < t_2 < \cdots < t_k$, $k \in \mathbb{N}$, by repeated conditioning and applying Eqn. (8). Further, as in Eqn. (12), we can obtain corresponding joint moments by differentiation. This in particular allows us to compute the auto-correlation and auto-covariance functions of the multivariate Hawkes process and its associated population process. More precisely, for any $t \geqslant 0$ and $\tau > 0$, we can compute the auto-correlation function by

$$R_{\boldsymbol{Q}}(t,\tau) = \mathbb{E}\big[\boldsymbol{Q}(t)\,\boldsymbol{Q}(t+\tau)^\top\big], \qquad R_{\boldsymbol{\lambda}}(t,\tau) = \mathbb{E}\big[\boldsymbol{\lambda}(t)\,\boldsymbol{\lambda}(t+\tau)^\top\big],$$

and the auto-covariance function by

$$\begin{aligned}
C_{\boldsymbol{Q}}(t,\tau) &= \mathbb{E}\big[\boldsymbol{Q}(t)\,\boldsymbol{Q}(t+\tau)^\top\big] - \mathbb{E}\big[\boldsymbol{Q}(t)\big]\mathbb{E}\big[\boldsymbol{Q}(t+\tau)\big]^\top,\\
C_{\boldsymbol{\lambda}}(t,\tau) &= \mathbb{E}\big[\boldsymbol{\lambda}(t)\,\boldsymbol{\lambda}(t+\tau)^\top\big] - \mathbb{E}\big[\boldsymbol{\lambda}(t)\big]\mathbb{E}\big[\boldsymbol{\lambda}(t+\tau)\big]^\top.
\end{aligned}$$

3.2. **Joint moments.** We proceed by exploiting the characterization of the joint transform $\zeta(t, \boldsymbol{s}, \boldsymbol{z})$, as given in Theorem 1, to derive a system of linear differential equations for the joint transient moments pertaining to $(\boldsymbol{\lambda}(t), \boldsymbol{Q}(t))$, as well as a system of linear (algebraic) equations for the corresponding stationary moments pertaining to $(\boldsymbol{\lambda}, \boldsymbol{Q})$, where

$$\begin{aligned}
\boldsymbol{\lambda} &= (\lambda_1, \ldots, \lambda_d) := \lim_{t\to\infty} (\lambda_1(t), \ldots, \lambda_d(t)),\\
\boldsymbol{Q} &= (Q_1, \ldots, Q_d) := \lim_{t\to\infty} (Q_1(t), \ldots, Q_d(t)).
\end{aligned} \tag{16}$$

As mentioned, Assumption 1 entails that $\boldsymbol{\lambda}$ exists, and if in addition all μ_i are positive, then $\boldsymbol{Q}$ is well defined as well.

We start by analyzing the joint transient moments via a system of linear differential equations. Let $n_{\lambda_i}, n_{Q_i} \in \mathbb{N}$ be such that $\sum_{i=1}^d n_{\lambda_i} = n_\lambda$ and $\sum_{i=1}^d n_{Q_i} = n_Q$. Consider, for any $t \in \mathbb{R}_+$, with $\boldsymbol{n_\lambda} = (n_{\lambda_1}, \ldots, n_{\lambda_d})$, $\boldsymbol{n_Q} = (n_{Q_1}, \ldots, n_{Q_d})$, the object of our interest:

$$\varphi_t(\boldsymbol{n_\lambda}, \boldsymbol{n_Q}) := \mathbb{E}\Big[\prod_{i=1}^d \lambda_i(t)^{n_{\lambda_i}} Q_i(t)^{n_{Q_i}}\Big]; \tag{17}$$

we call $\varphi_t(\boldsymbol{n_\lambda}, \boldsymbol{n_Q})$ *joint transient moments of total order* n_λ *and* n_Q. To make sure the objects that we consider are well-defined, we throughout assume that, for any $j \in [d]$,

$$\mathbb{E}\Big[\prod_{i=1}^d B_{ij}^{n_{\lambda_i}}\Big] < \infty. \tag{18}$$

By generalizing the approach of [18] to the multivariate setting, we obtain a vector-valued ODE (ordinary differential equation) to derive the joint transient moments (17). In this derivation, a crucial role is played by an intermediate step in the proof of Theorem 1, as given in Appendix A. In this intermediate step, summarized in Eqn. (58), the following PDE has been derived:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\Big[e^{-\boldsymbol{s}^\top\boldsymbol{\lambda}(t)}\prod_{n=1}^d z_n^{Q_n(t)}\Big] - \sum_{j=1}^d\big(\alpha_j s_j + z_j\beta_j(\boldsymbol{s})-1\big)\mathbb{E}\big[\lambda_j(t)e^{-\boldsymbol{s}^\top\boldsymbol{\lambda}(t)}\prod_{n=1}^d z_n^{Q_n(t)}\big] \qquad (19)\\ +\sum_{j=1}^d \mu_j(z_j-1)\mathbb{E}\big[Q_j(t)e^{-\boldsymbol{s}^\top\boldsymbol{\lambda}(t)}\prod_{n=1}^d z_n^{Q_n(t)-\mathbf{1}_{\{n=j\}}}\big] = -\sum_{j=1}^d \alpha_j s_j\overline{\lambda}_j\mathbb{E}\big[e^{-\boldsymbol{s}^\top\boldsymbol{\lambda}(t)}\prod_{n=1}^d z_n^{Q_n(t)}\big],\end{aligned}$$

where we rewrote each of the terms appearing in Eqn. (58) using the definition of $\zeta(t,\boldsymbol{s},\boldsymbol{z})$. The next step is to repeatedly differentiate this PDE: differentiate $n_{\lambda_1},\dots,n_{\lambda_d}$ times with respect to $s_1,\dots,s_d$, respectively and substitute $\boldsymbol{s}=\mathbf{0}$, and then differentiate $n_{Q_1},\dots,n_{Q_d}$ times with respect to $z_1,\dots,z_d$ respectively and substitute $\boldsymbol{z}=\mathbf{1}$. To make our notation concise, we introduce the 'reduced version' of $\varphi_t(\boldsymbol{n_Q},\boldsymbol{n_\lambda})$:

$$\psi_t(\boldsymbol{n_\lambda},\boldsymbol{n_Q}) := \mathbb{E}\Big[\prod_{i=1}^d \lambda_i(t)^{n_{\lambda_i}}Q_i(t)^{[n_{Q_i}]}\Big], \qquad (20)$$

where evidently $\psi_t(\boldsymbol{n_Q},\boldsymbol{n_\lambda})$ can be expressed in objects of the type $\varphi_t(\boldsymbol{n_Q},\boldsymbol{n_\lambda})$ as given in Eqn. (17), and vice versa. By elementary algebraic operations, we obtain, with $\sum_{=1}^d m_i = m$,

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\psi_t(\boldsymbol{n_\lambda},\boldsymbol{n_Q}) + \sum_{j=1}^d\big(n_{\lambda_j}(\alpha_j-\mathbb{E}[B_{jj}]) + n_{Q_j}\mu_j\big)\psi_t(\boldsymbol{n_\lambda},\boldsymbol{n_Q})\\ &= \sum_{j=1}^d\sum_{\substack{i=1\\i\neq j}}^d n_{\lambda_i}\mathbb{E}[B_{ij}]\psi_t(\boldsymbol{n_\lambda}-\boldsymbol{e}_i+\boldsymbol{e}_j,\boldsymbol{n_Q}) + \sum_{j=1}^d n_{Q_j}\psi_t(\boldsymbol{n_\lambda}+\boldsymbol{e_j},\boldsymbol{n_Q}-\boldsymbol{e}_j) \qquad (21)\\ &\quad+\sum_{j=1}^d\alpha_j\overline{\lambda}_j n_{\lambda_j}\psi_t(\boldsymbol{n_\lambda}-\boldsymbol{e}_j,\boldsymbol{n_Q}) + \sum_{i=1}^d\sum_{j=1}^d n_{\lambda_i}n_{Q_j}\mathbb{E}[B_{ij}]\psi_t(\boldsymbol{n_\lambda}-\boldsymbol{e}_i+\boldsymbol{e}_j,\boldsymbol{n_Q}-\boldsymbol{e}_j)\\ &\quad+\sum_{j=1}^d\sum_{m_1=0}^{n_{\lambda_1}}\cdots\sum_{m_d=0}^{n_{\lambda_d}}\mathbf{1}_{\{m\leqslant n_\lambda-2\}}\prod_{k=1}^d\binom{n_{\lambda_k}}{m_k}\Big\{n_{Q_j}\prod_{i=1}^d\mathbb{E}\big[B_{ij}^{n_{\lambda_i}-m_i}\big]\psi_t(\boldsymbol{m}+\boldsymbol{e}_j,\boldsymbol{n_Q}-\boldsymbol{e}_j)\\ &\quad+\prod_{i=1}^d\mathbb{E}\big[B_{ij}^{n_{\lambda_i}-m_i}\big]\psi_t(\boldsymbol{m}+\boldsymbol{e}_j,\boldsymbol{n_Q})\Big\};\end{aligned}$$

see Appendix B.2 for details. The key observation is that Eqn. (21) provides us with a relation involving joint (reduced) transient moments of $(\boldsymbol{\lambda}(t),\boldsymbol{Q}(t))$ of total order n_λ and n_Q, expressed in terms of their counterparts of at most the same total order, so that the system can be solved. The resulting linear vector-valued ODE thus enables the computation of the joint transient moments of total order n_λ and n_Q, thus generalizing Eqn. (3.9) in [18].

After having dealt with the joint transient moments, we now focus on the stationary counterpart of $\psi_t(\boldsymbol{n_\lambda}, \boldsymbol{n_Q})$, i.e., the *joint reduced stationary moments of total order* n_λ *and* n_Q:

$$\psi(\boldsymbol{n_\lambda}, \boldsymbol{n_Q}) := \lim_{t\to\infty} \psi_t(\boldsymbol{n_\lambda}, \boldsymbol{n_Q}) = \mathbb{E}\Big[\prod_{i=1}^{d} \lambda_i^{n_{\lambda_i}} Q_i^{[n_{Q_i}]}\Big]. \tag{22}$$

In this case, we obtain a system of algebraic equations, to be interpreted as the stationary version of Eqn. (21). Indeed, it involves joint (reduced) stationary moments of total order n_λ and n_Q, expressed in terms of their counterparts of at most the same total order:

$$\begin{aligned}
&\sum_{j=1}^{d} \big(n_{\lambda_j}(\alpha_j - \mathbb{E}\big[B_{jj}\big]) + n_{Q_j}\mu_j\big)\, \psi(\boldsymbol{n_\lambda}, \boldsymbol{n_Q}) \\
&= \sum_{j=1}^{d} \sum_{\substack{i=1\\ i\neq j}}^{d} n_{\lambda_i}\mathbb{E}\big[B_{ij}\big]\psi(\boldsymbol{n_\lambda} - \boldsymbol{e}_i + \boldsymbol{e}_j, \boldsymbol{n_Q}) + \sum_{j=1}^{d} n_{Q_j}\psi(\boldsymbol{n_\lambda} + \boldsymbol{e_j}, \boldsymbol{n_Q} - \boldsymbol{e}_j) \\
&\quad + \sum_{j=1}^{d} \alpha_j\overline{\lambda}_j n_{\lambda_j}\psi(\boldsymbol{n_\lambda} - \boldsymbol{e}_j, \boldsymbol{n_Q}) + \sum_{i=1}^{d}\sum_{j=1}^{d} n_{\lambda_i} n_{Q_j}\mathbb{E}\big[B_{ij}\big]\psi(\boldsymbol{n_\lambda} - \boldsymbol{e}_i + \boldsymbol{e}_j, \boldsymbol{n_Q} - \boldsymbol{e}_j) \\
&\quad + \sum_{j=1}^{d}\sum_{m_1=0}^{n_{\lambda_1}} \cdots \sum_{m_d=0}^{n_{\lambda_d}} \mathbf{1}_{\{m \leqslant n_\lambda - 2\}} \prod_{k=1}^{d} \binom{n_{\lambda_k}}{m_k} \Big\{ n_{Q_j} \prod_{i=1}^{d} \mathbb{E}\big[B_{ij}^{n_{\lambda_i}-m_i}\big]\psi(\boldsymbol{m} + \boldsymbol{e}_j, \boldsymbol{n_Q} - \boldsymbol{e}_j) \\
&\quad + \prod_{i=1}^{d} \mathbb{E}\big[B_{ij}^{n_{\lambda_i}-m_i}\big]\psi(\boldsymbol{m} + \boldsymbol{e}_j, \boldsymbol{n_Q})\Big\}.
\end{aligned} \tag{23}$$

In Appendix C we present an illustration concerning moments of order $n \in \{1, 2\}$.

4. Recursive Procedure: Bivariate Setting

In this section, we specifically consider the bivariate setting ($d = 2$), and focus on the structure behind the joint moments of arbitrary order $n \in \mathbb{N}$. The presented method can be extended to higher dimensions $d \in \mathbb{N}$, at the cost of heavier notation and more intricate objects. As such, this section serves as a proof of principle on how the underlying recursive structure can be exploited. We construct, based on the results obtained in the previous sections, a recursive procedure to compute the joint transient moments $\psi_t(\boldsymbol{n_\lambda}, \boldsymbol{n_Q})$ as well as the joint stationary moments $\psi(\boldsymbol{n_\lambda}, \boldsymbol{n_Q})$. To make the analysis as transparent as possible, we express the main objects in vector/matrix-form. As it turns out, there is a strong similarity between the structure of the algorithm to evaluate the transient moments on one hand, and its counterpart for the stationary moments on the other hand. In the sequel we let n be the *total order* of the joint moments, i.e., $n = n_Q + n_\lambda = n_{Q_1} + n_{Q_2} + n_{\lambda_1} + n_{\lambda_2}$. We first rewrite the coupled equations (21) (transient case) and (23) (stationary case) in vector-matrix form. We then use these to set up a procedure to compute the corresponding moments.

4.1. **Transient moments.** We construct a recursive procedure to compute transient joint moments $\psi_t(\boldsymbol{n_\lambda}, \boldsymbol{n_Q})$ by introducing properly defined vector- and matrix-valued objects, such that we can exploit the ODE in Eqn. (21). As we have $d = 2$, the objective is to compute

$$\psi_t((n_{Q_1}, n_{Q_2}), (n_{\lambda_1}, n_{\lambda_2})) = \mathbb{E}\Big[\lambda_1(t)^{n_{\lambda_1}} \lambda_2(t)^{n_{\lambda_2}} Q_1(t)^{[n_{Q_1}]} Q_2(t)^{[n_{Q_2}]}\Big].$$

The multivariate setting has the intrinsic complication that the number of combinations of possible joint moments increases rapidly in n and d; already in this bivariate setting, there are many possible combinations of joint moments of order n. To collect all joint moments $\psi_t((n_{Q_1}, n_{Q_2}), (n_{\lambda_1}, n_{\lambda_2}))$ in a single vector, we need to specify an ordering of the different moments. To that end, we introduce the *stacked* vector

$$\boldsymbol{\Psi}_t^{(n)} := \Big(\boldsymbol{\Psi}_t^{(0,n)}, \boldsymbol{\Psi}_t^{(1,n-1)}, \ldots, \boldsymbol{\Psi}_t^{(n,0)}\Big)^\top, \tag{24}$$

where, for each $k \in \{0, 1, \ldots, n\}$, the vector $\boldsymbol{\Psi}_t^{(k,n-k)}$ exhaustively contains all combinations of joint moments such that $n_{Q_1} + n_{Q_2} = k$ and $n_{\lambda_1} + n_{\lambda_2} = n - k$. For instance, for $k = 0$ and $k = n$ we respectively have

$$\begin{aligned}
\boldsymbol{\Psi}_t^{(0,n)} &= \Big(\psi_t(\mathbf{0}, (n, 0)), \psi_t(\mathbf{0}, (n-1, 1)), \ldots, \psi_t(\mathbf{0}, (0, n))\Big)^\top \\
&= \Big(\mathbb{E}\big[\lambda_1(t)^n\big], \mathbb{E}\big[\lambda_1(t)^{n-1}\lambda_2(t)\big], \ldots, \mathbb{E}\big[\lambda_2(t)^n\big]\Big)^\top; \\
\boldsymbol{\Psi}_t^{(n,0)} &= \Big(\psi_t((n, 0), \mathbf{0}), \psi_t((n-1, 1), \mathbf{0}), \ldots, \psi_t((0, n), \mathbf{0})\Big)^\top \\
&= \Big(\mathbb{E}\big[Q_1(t)^{[n]}\big], \mathbb{E}\big[Q_1(t)^{[n-1]}Q_2(t)\big], \ldots, \mathbb{E}\big[Q_2(t)^{[n]}\big]\Big)^\top.
\end{aligned}$$

The cases corresponding with $k \in \{1, \ldots, n-1\}$ are notationally considerably more burdensome since one has to include all possible combinations of order k as well as $n - k$.

For concrete examples of the stacked vector in (24) for orders $n = 1, 2, 3$, see Appendix D.1. It is readily verified that for general d the dimension of $\boldsymbol{\Psi}_t^{(n)}$ equals

$$\mathfrak{D}(d, n) := \sum_{k=1}^{n} \binom{2d}{k} \binom{n-1}{k-1}, \tag{25}$$

where the $2d$ is due to the fact that we include moments of both $\boldsymbol{Q}(t)$ and $\boldsymbol{\lambda}(t)$. In the $d = 2$ case considered in this section, we thus have that the size of $\boldsymbol{\Psi}_t^{(n)}$ is $\mathfrak{D}(2, n)$.

By Eqn. (21), the stacked vector $\boldsymbol{\Psi}_t^{(n)}$ satisfies a vector-valued ODE:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Psi}_t^{(n)} = \boldsymbol{M}\boldsymbol{\Psi}_t^{(n)} + \boldsymbol{L}\big(\boldsymbol{\Psi}_t^{(1)}, \ldots, \boldsymbol{\Psi}_t^{(n-1)}\big)^\top, \tag{26}$$

for certain matrices $\boldsymbol{M}$ and $\boldsymbol{L}$ of appropriate dimension. Here, the matrix $\boldsymbol{M}$ is of dimension $\mathfrak{D}(2, n) \times \mathfrak{D}(2, n)$, and $\boldsymbol{L}$ of dimension $\mathfrak{D}(2, n) \times \overline{\mathfrak{D}}(2, n)$, where

$$\overline{\mathfrak{D}}(2, n) := \sum_{m=1}^{n-1} \mathfrak{D}(2, m). \tag{27}$$

As a next step, we identify blocks of $\boldsymbol{M}$ that correspond to subvectors $\boldsymbol{\Psi}_t^{(k,n-k)}$ of the stacked vector $\boldsymbol{\Psi}_t^{(n)}$. Upon inspecting (21) we observe that, when considering in (26) the differential equations that correspond to $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Psi}_t^{(k,n-k)}$, in the right-hand side only $\boldsymbol{\Psi}_t^{(k,n-k)}$ and $\boldsymbol{\Psi}_t^{(k-1,n-k+1)}$

appear, besides a linear combination of objects of lower total order (i.e., $\boldsymbol{\Psi}_t^{(1)},\dots,\boldsymbol{\Psi}_t^{(n-1)}$). As a consequence, we can write

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Psi}_t^{(k,n-k)} = \boldsymbol{M}^{(k,n-k)}\boldsymbol{\Psi}_t^{(k,n-k)} + \boldsymbol{K}^{(k,n-k)}\boldsymbol{\Psi}_t^{(k-1,n-k+1)} + \boldsymbol{L}^{(k,n-k)}\big(\boldsymbol{\Psi}_t^{(1)},\dots,\boldsymbol{\Psi}_t^{(n-1)}\big)^\top, \quad (28)$$

for appropriately chosen matrices $\boldsymbol{M}^{(k,n-k)}$, $\boldsymbol{K}^{(k,n-k)}$, and $\boldsymbol{L}^{(k,n-k)}$, where we set $\boldsymbol{K}^{(k,n-k)}\equiv 0$ when $k=0$. Eqn. (28) thus reveals a recursive procedure to compute $\boldsymbol{\Psi}_t^{(n)}$, where in the n-th iteration a non-homogeneous linear system of ODEs has to be solved, with $\boldsymbol{\Psi}_t^{(1)},\dots,\boldsymbol{\Psi}_t^{(n-1)}$, as derived in the previous steps, appearing in the non-homogeneous part.

We proceed by introducing the notation needed to set up the recursive procedure. A *tridiagonal* matrix in $\mathbb{R}^{n\times n}$ is a matrix with elements on the main diagonal, the first diagonal above and below the main diagonal only, for which we use the notation, with $\boldsymbol{a},\boldsymbol{c}\in\mathbb{R}^{n-1}$ and $\boldsymbol{d}\in\mathbb{R}^n$,

$$\operatorname{tridiag}\big(\boldsymbol{a},\boldsymbol{d},\boldsymbol{c}\big) := \begin{bmatrix} d_1 & c_1 & 0 & \cdots & 0 \\ a_1 & d_2 & c_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & a_{n-2} & d_{n-1} & c_{n-1} \\ 0 & 0 & 0 & a_{n-1} & d_n \end{bmatrix}.$$

Given the vectors $\boldsymbol{n_Q}=(n_{Q_1},n_{Q_2})$ and $\boldsymbol{n_\lambda}=(n_{\lambda_1},n_{\lambda_2})$, we set

$$\begin{aligned} v(\boldsymbol{n_Q},\boldsymbol{n_\lambda}) &:= -n_{\lambda_1}\overline{\alpha}_1 - n_{Q_1}\mu_1 - n_{\lambda_2}\overline{\alpha}_2 - n_{Q_2}\mu_2, \\ \boldsymbol{w}(\boldsymbol{n_Q},n) &:= \big(v(\boldsymbol{n_Q},(n,0)), v(\boldsymbol{n_Q},(n-1,1)),\dots,v(\boldsymbol{n_Q},(1,n-1)),v(\boldsymbol{n_Q},(0,n))\big)^\top, \end{aligned}$$

where $\overline{\alpha}_i := \alpha_i - \mathbb{E}[B_{ii}]$ for $i=1,2$, corresponding to the left-hand side of Eqn. (21). With the vectors $\boldsymbol{1}_{(n)} := (1,2,\dots,n)$ and $\boldsymbol{1}^{(n)} := (n,n-1,\dots,1)$, and with e_j the unit vector with 1 on the $j-$th component, we finally define the matrix, for each $k\in\{0,1,\dots,n-1,n\}$,

$$\boldsymbol{M}^{(k,n-k)} := \bigoplus_{m=0}^{k}\operatorname{tridiag}\big(\boldsymbol{1}_{(n-k)}\mathbb{E}[B_{21}], \boldsymbol{w}\big((n-m)\boldsymbol{e}_1+m\boldsymbol{e}_2,n-k\big),\boldsymbol{1}^{(n-k)}\mathbb{E}[B_{12}]\big), \quad (29)$$

where $\bigoplus$ denotes the direct sum for matrices. For concrete examples of the matrix $\boldsymbol{M}^{(k,n-k)}$ and how it appears in the ODE in (28), see Appendix D. We can now present our algorithm to compute the transient moments.

Algorithm 1. *Fix $n\in\mathbb{N}$, and suppose we know $\boldsymbol{\Psi}_t^{(1)},\dots,\boldsymbol{\Psi}_t^{(n-1)}$. Then $\boldsymbol{\Psi}_t^{(n)}$ can be found from Eqn.* (26)*, as follows:*

Step 0: The vector $\boldsymbol{\Psi}_t^{(0,n)}$ satisfies the vector-valued ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Psi}_t^{(0,n)} = \boldsymbol{M}^{(0,n)}\boldsymbol{\Psi}_t^{(0,n)} + \boldsymbol{L}^{(0,n)}\big(\boldsymbol{\Psi}_t^{(1)},\dots,\boldsymbol{\Psi}_t^{(n-1)}\big), \quad (30)$$

where $\boldsymbol{L}^{(0,n)}$ follows from Eqn. (21)*, with initial condition $\boldsymbol{\Psi}_0^{(0,n)}$ determined by $\boldsymbol{Q}(0)$ and $\boldsymbol{\lambda}(0)$.*

Step k: For any $k \in \{1, 2, \ldots, n-1, n\}$, $\boldsymbol{\Psi}_t^{(k,n-k)}$ satisfies the vector-valued ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Psi}_t^{(k,n-k)} = \boldsymbol{M}^{(k,n-k)}\boldsymbol{\Psi}_t^{(k,n-k)} + \boldsymbol{K}^{(k,n-k)}\boldsymbol{\Psi}_t^{(k-1,n-k+1)} \qquad (31)$$
$$+ \boldsymbol{L}^{(k,n-k)}\big(\boldsymbol{\Psi}_t^{(1)}, \ldots, \boldsymbol{\Psi}_t^{(n-1)}\big),$$

where $\boldsymbol{K}^{(k,n-k)}$ and $\boldsymbol{L}^{(k,n-k)}$ follow from Eqn. (21), with initial condition $\boldsymbol{\Psi}_0^{(k,n-k)}$.

Clearly, Eqn. (21) uniquely defines the matrices $\boldsymbol{K}^{(k,n-k)}$ and $\boldsymbol{L}^{(k,n-k)}$ needed in the above algorithm. However, their explicit definition would require objects that are even more notationally involved. In Appendix D we show that for moments of orders $n = 1$ and $n = 2$, we can still explicitly write down the matrix for the stacked vector ODE: we combine the blocks of matrices into 4×4 and 10×10 matrices $\boldsymbol{M}$, respectively, and also construct the corresponding matrix $\boldsymbol{L}$. However, for order $n = 3, 4, \ldots$, we would need very large matrices which are cumbersome to write down explicitly.

Due to the direct sum structure of $\boldsymbol{M}^{(k,n-k)}$, it consists of blocks. This allows us to decompose the k-th step in the algorithm into smaller steps, by considering the parts of the vector $\boldsymbol{\Psi}_t^{(k,n-k)}$ associated with the individual blocks of the matrix. The solution to the ODEs in Algorithm 1 can be given in terms of matrix exponentials, as follows. This result follows by observing that computing $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Psi}_t^{(k,n-k)}$ in Eqn. (32) immediately yields (31) by inspection and Leibniz' integral rule.

Proposition 1. *For fixed $t \in \mathbb{R}_+$, $n \in \mathbb{N}$ and $k = 0, 1, \ldots, n$, the solution for the vector-valued ODE for $\boldsymbol{\Psi}_t^{(k,n-k)}$ in Eqn. (31) is given by*

$$\boldsymbol{\Psi}_t^{(k,n-k)} = e^{t\boldsymbol{M}^{(k,n-k)}}\boldsymbol{\Psi}_0^{(k,n-k)} \qquad (32)$$
$$+ \int_0^t e^{(t-s)\boldsymbol{M}^{(k,n-k)}}\Big(\boldsymbol{K}^{(k,n-k)}\boldsymbol{\Psi}_s^{(k-1,n-k+1)} + \boldsymbol{L}^{(k,n-k)}\big(\boldsymbol{\Psi}_s^{(1)}, \ldots, \boldsymbol{\Psi}_s^{(n-1)}\big)\Big)\,\mathrm{d}s.$$

4.2. **Stationary moments.** In this subsection, we focus on, with $n = n_Q + n_\lambda$,

$$\psi((n_{Q_1}, n_{Q_2}), (n_{\lambda_1}, n_{\lambda_2})) = \mathbb{E}\Big[\lambda_1^{n_{\lambda_1}}\lambda_2^{n_{\lambda_2}}Q_1^{[n_{Q_1}]}Q_2^{[n_{Q_2}]}\Big]. \qquad (33)$$

By exploiting Eqn. (23), we develop a recursive procedure similar to the one for the transient moments. The central object of study is

$$\boldsymbol{\Psi}^{(n)} := \lim_{t\to\infty}\boldsymbol{\Psi}_t^{(n)} = \Big(\boldsymbol{\Psi}^{(0,n)}, \boldsymbol{\Psi}^{(1,n-1)}, \ldots, \boldsymbol{\Psi}^{(n,0)}\Big)^\top. \qquad (34)$$

For the following recursive procedure, we use the notation introduced in Section 4.1; recall in particular the matrices defined in Eqn. (29). Observe the strong similarity between the Algorithms 1 and 2, in the sense that the underlying recursive structures fully match.

Algorithm 2. *Fix $n \in \mathbb{N}$, and suppose we know $\boldsymbol{\Psi}^{(1)}, \ldots, \boldsymbol{\Psi}^{(n-1)}$. Then $\boldsymbol{\Psi}^{(n)}$ can be computed as follows:*

Step 0: The vector $\boldsymbol{\Psi}^{(0,n)}$ satisfies the linear equation

$$0 = \boldsymbol{M}^{(0,n)}\,\boldsymbol{\Psi}^{(0,n)} + \boldsymbol{L}^{(0,n)}\big(\boldsymbol{\Psi}^{(1)}, \ldots, \boldsymbol{\Psi}^{(n-1)}\big)^\top, \qquad (35)$$

where $\boldsymbol{L}^{(0,n)}$ *follows from Eqn.* (23)*.*

Step k: For any $k \in \{1, 2, \ldots, n-1, n\}$*,* $\boldsymbol{\Psi}^{(k,n-k)}$ *satisfies the linear equations*

$$0 = \boldsymbol{M}^{(k,n-k)}\,\boldsymbol{\Psi}^{(k,n-k)} + \boldsymbol{K}^{(k,n-k)}\boldsymbol{\Psi}^{(k-1,n-k+1)} + \boldsymbol{L}^{(k,n-k)}\big(\boldsymbol{\Psi}^{(1)}, \ldots, \boldsymbol{\Psi}^{(n-1)}\big)^{\top}. \tag{36}$$

where $\boldsymbol{K}^{(k,n-k)}$ *and* $\boldsymbol{L}^{(k,n-k)}$ *follow from Eqn.* (23)*.*

The solution to the linear equations in Algorithm 2 is given in the following proposition. Its proof follows immediately from solving Eqn. (36).

Proposition 2. *For* $n \in \mathbb{N}$ *and* $k = 0, 1, \ldots, n-1, n$*, the solution for the linear equation for* $\boldsymbol{\Psi}^{(k,n-k)}$ *in Eqn.* (36) *is given by*

$$\boldsymbol{\Psi}^{(k,n-k)} = -\Big(\boldsymbol{M}^{(k,n-k)}\Big)^{-1}\Big\{\boldsymbol{K}^{(k,n-k)}\boldsymbol{\Psi}^{(k-1,n-k+1)} + \boldsymbol{L}^{(k,n-k)}\big(\boldsymbol{\Psi}^{(1)}, \ldots, \boldsymbol{\Psi}^{(n-1)}\big)^{\top}\Big\},$$

5. Nested Block Matrices: Bivariate Setting

In this section, we investigate the nested structure of the matrices associated with the ODEs of the moments more thoroughly, again in the bivariate setting $d = 2$ (but, as before, extension to higher d is in principle possible), to develop an efficient computational method for evaluating the transient and stationary moments. It turns out that one can find a nested sequence of well-behaved matrices that describe the relations between the moments, which facilitates a substantial reduction of the required computational effort. The approach followed in this section can be seen as a bivariate version of Section 3.2 in [11], providing the structure of ODEs associated with the transient moments using lower triangular matrices with scalar entries. The key difference, however, is that in our case they are replaced by block lower triangular matrices, containing matrix entries.

We first revisit the transient moments of $\boldsymbol{\lambda}(t)$ for a fixed $t \in \mathbb{R}_+$ so as to illustrate the nested structure of the matrices. After that, we consider the joint transient moments of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$, which has a similar but more complex nested structure. To motivate our analysis, consider the ODEs associated with the vectors $\boldsymbol{\Psi}_t^{(0,1)}$ and $\boldsymbol{\Psi}_t^{(0,2)}$ containing the first and second order moments of $\boldsymbol{\lambda}(t)$ (see Eqns. (78) and (80) in Appendix D.1). Observe that in stacked form, they can be represented in a block lower triangular matrix structure:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\boldsymbol{\Psi}_t^{(0,1)}\\ \boldsymbol{\Psi}_t^{(0,2)}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A}_1^{2\times 2} & \boldsymbol{0}_2^{2\times 3}\\ \boldsymbol{D}_2^{3\times 2} & \boldsymbol{C}_2^{3\times 3}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Psi}_t^{(0,1)}\\ \boldsymbol{\Psi}_t^{(0,2)}\end{bmatrix} + \begin{bmatrix}\boldsymbol{b}^{2\times 1}\\ \boldsymbol{0}^{3\times 1}\end{bmatrix}, \tag{37}$$

with $\boldsymbol{A}_1^{2\times 2} = \boldsymbol{M}^{(0,1)}$, $\boldsymbol{C}_2^{3\times 3} = \boldsymbol{M}^{(0,2)}$, $\boldsymbol{D}_2^{3\times 2} = \boldsymbol{L}^{(0,2)}$, where the superscripts denote the dimensionality of the matrices, and where $\boldsymbol{M}^{(0,1)}$ and $\boldsymbol{M}^{(0,2)}$ are defined in Eqn. (29) and $\boldsymbol{L}^{(0,2)}$ in Eqn. (80). The notation $\boldsymbol{0}^{k\times l} \in \mathbb{R}^{k\times l}$ represents the all-zeros matrix and $\boldsymbol{b}^{2\times 1} = (\alpha_1\overline{\lambda}_1, \alpha_2\overline{\lambda}_2)^{\top}$. Observe that this stacked form contains the previously defined matrices as blocks. A similar form occurs for higher orders n of the stacked vector $(\boldsymbol{\Psi}_t^{(0,1)}, \ldots, \boldsymbol{\Psi}_t^{(0,n)})^{\top}$, containing mixed moments of $\boldsymbol{\lambda}(t)$ up to order n.

Careful inspection of previous results reveals a *nested sequence* of *block lower triangular matrices.* Let $m_n := n+1$ and define $\mathfrak{m}_n := m_1 + \cdots + m_n$. Then, consider a nested sequence of block lower triangular matrices $\{\boldsymbol{A}_n^{\mathfrak{m}_n\times\mathfrak{m}_n}\}_{n\in\mathbb{N}}$ defined by

$$\boldsymbol{A}_n^{\mathfrak{m}_n\times\mathfrak{m}_n} = \begin{bmatrix} \boldsymbol{A}_{n-1}^{\mathfrak{m}_{n-1}\times\mathfrak{m}_{n-1}} & \boldsymbol{0}_n^{\mathfrak{m}_{n-1}\times m_n} \\ \boldsymbol{D}_n^{m_n\times\mathfrak{m}_{n-1}} & \boldsymbol{C}_n^{m_n\times m_n} \end{bmatrix}, \tag{38}$$

where $\boldsymbol{A}_1^{2\times2} \equiv \boldsymbol{C}_1^{2\times2} = \boldsymbol{M}^{(0,1)}$, $\boldsymbol{C}_n^{m_n\times m_n} = \boldsymbol{M}^{(0,n)}$ and $\boldsymbol{D}_n^{m_n\times\mathfrak{m}_{n-1}} = \boldsymbol{L}^{(0,n)}$. Clearly, the first term on the right-hand side of (37) occurs as a special case of (38) when $n=2$. Recall that we know the structure of the matrices $\boldsymbol{M}^{(0,n)}$ as given in Eqn. (29). In Appendix D.4, we give some further details on the structure of $\boldsymbol{L}^{(0,n)}$ for the case $n=3$. The sequence of matrices $\{\boldsymbol{A}_n^{\mathfrak{m}_n\times\mathfrak{m}_n}\}_{n\in\mathbb{N}}$, as defined in (38), has been chosen such that

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \boldsymbol{\Psi}_t^{(0,1)} \\ \vdots \\ \boldsymbol{\Psi}_t^{(0,n)} \end{bmatrix} = \boldsymbol{A}_n^{\mathfrak{m}_n\times\mathfrak{m}_n}\begin{bmatrix} \boldsymbol{\Psi}_t^{(0,1)} \\ \vdots \\ \boldsymbol{\Psi}_t^{(0,n)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{b}^{2\times1} \\ \boldsymbol{0}^{(\mathfrak{m}_n-m_1)\times1} \end{bmatrix}, \tag{39}$$

with initial condition $(\boldsymbol{\Psi}_0^{(0,1)},\ldots,\boldsymbol{\Psi}_0^{(0,n)})^\top$.

The following proposition, providing an explicit expression for $(\boldsymbol{\Psi}_t^{(0,1)},\ldots,\boldsymbol{\Psi}_t^{(0,n)})^\top$, follows directly by noting that taking the time derivative of Eqn. (40) immediately yields Eqn. (39).

Proposition 3. *If $\boldsymbol{C}_i^{m_i\times m_i}$ is invertible for all $i\in\{1,\ldots,n\}$, then*

$$\begin{bmatrix} \boldsymbol{\Psi}_t^{(0,1)} \\ \vdots \\ \boldsymbol{\Psi}_t^{(0,n)} \end{bmatrix} = e^{\boldsymbol{A}_n^{\mathfrak{m}_n\times\mathfrak{m}_n}t}\begin{bmatrix} \boldsymbol{\Psi}_0^{(0,1)} \\ \vdots \\ \boldsymbol{\Psi}_0^{(0,n)} \end{bmatrix} - \left(\boldsymbol{A}_n^{\mathfrak{m}_n\times\mathfrak{m}_n}\right)^{-1}\left(\boldsymbol{I}^{\mathfrak{m}_n\times\mathfrak{m}_n} - e^{\boldsymbol{A}_n^{\mathfrak{m}_n\times\mathfrak{m}_n}t}\right)\begin{bmatrix} \boldsymbol{b}^{2\times1} \\ \boldsymbol{0}^{(\mathfrak{m}_n-m_1)\times1} \end{bmatrix}. \tag{40}$$

Proposition 3 allows for the simultaneous computation of the first n transient moments of $\boldsymbol{\lambda}(t) = (\lambda_1(t),\lambda_2(t))^\top$, but it requires the computation of the matrix exponential and the inverse of $\boldsymbol{A}_n^{\mathfrak{m}_n\times\mathfrak{m}_n}$.

This idea can be extended to the joint transient moments of $(\boldsymbol{Q}(t),\boldsymbol{\lambda}(t))$. As before, we show the details of the first and second order, and point out how this extends to higher order moments. For the first order moments, close inspection of the associated ODEs (given in Eqns. (78) and (79) in Appendix D.1) yields that

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Psi}_t^{(1)} = \boldsymbol{F}_1^{4\times4}\boldsymbol{\Psi}_t^{(1)} + \begin{bmatrix} \boldsymbol{b}^{2\times1} \\ \boldsymbol{0}^{2\times1} \end{bmatrix}, \text{ with } \boldsymbol{F}_1^{4\times4} = \begin{bmatrix} \boldsymbol{M}^{(0,1)} & \boldsymbol{0}^{2\times2} \\ \boldsymbol{I}^{2\times2} & \boldsymbol{M}^{(1,0)} \end{bmatrix}, \quad \boldsymbol{b}_1^{2\times1} = \begin{bmatrix} \alpha_1\overline{\lambda}_1 \\ \alpha_2\overline{\lambda}_2 \end{bmatrix}. \tag{41}$$

As before, the superscripts denote the dimensionality of the matrices, with $\boldsymbol{I}^{k\times l}\in\mathbb{R}^{k\times l}$ the identity matrix. Note the block lower triangular shape of $\boldsymbol{F}_1^{4\times4}$.

For the second order moments, we can infer from the associated ODEs (given in Eqns. (80), (81) and (82) in Appendix D.1), in combination with Eqn. (41), that

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \boldsymbol{\Psi}_t^{(1)} \\ \boldsymbol{\Psi}_t^{(2)} \end{bmatrix} = \boldsymbol{F}_2^{14\times14}\begin{bmatrix} \boldsymbol{\Psi}_t^{(1)} \\ \boldsymbol{\Psi}_t^{(2)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{b}_1^{2\times1} \\ \boldsymbol{0}^{12\times1} \end{bmatrix}, \text{ with } \boldsymbol{F}_2^{14\times14} = \begin{bmatrix} \boldsymbol{F}_1^{4\times4} & \boldsymbol{0}^{4\times10} \\ \boldsymbol{G}_2^{10\times4} & \boldsymbol{H}_2^{10\times10}, \end{bmatrix}; \tag{42}$$

here $\boldsymbol{F}_2^{14\times 14}$ is a lower triangular matrix with the matrices contained in it defined by

$$\boldsymbol{G}_2^{10\times 4} = \begin{bmatrix} \boldsymbol{L}^{(0,2)} & \boldsymbol{0}^{3\times 2} \\ \boldsymbol{L}_\lambda^{(1,1)} & \boldsymbol{L}_Q^{(1,1)} \\ \boldsymbol{0}^{3\times 2} & \boldsymbol{0}^{3\times 2} \end{bmatrix}, \quad \boldsymbol{H}_2^{10\times 10} = \begin{bmatrix} \boldsymbol{M}^{(0,2)} & \boldsymbol{0}^{3\times 4} & \boldsymbol{0}^{3\times 3} \\ \boldsymbol{K}^{(1,1)} & \boldsymbol{M}^{(1,1)} & \boldsymbol{0}^{3\times 3} \\ \boldsymbol{0}^{3\times 3} & \boldsymbol{K}^{(2,0)} & \boldsymbol{M}^{(2,0)} \end{bmatrix},$$

$$\boldsymbol{L}_\lambda^{(1,1)} = \begin{bmatrix} \mathbb{E}[B_{11}] & 0 \\ \mathbb{E}[B_{21}] & 0 \\ 0 & \mathbb{E}[B_{12}] \\ 0 & \mathbb{E}[B_{22}] \end{bmatrix}, \quad \boldsymbol{L}_Q^{(1,1)} = \begin{bmatrix} \alpha_1\overline{\lambda}_1 & 0 \\ \alpha_2\overline{\lambda}_2 & 0 \\ 0 & \alpha_1\overline{\lambda}_1 \\ 0 & \alpha_2\overline{\lambda}_2 \end{bmatrix},$$

where $\boldsymbol{L}^{(0,2)}$ and the matrices $\boldsymbol{K}^{(1,1)}$ and $\boldsymbol{K}^{(2,0)}$ are known, see Appendix D.1.

Continuing in this fashion we can consider vectors of arbitrary length $n \in \mathbb{N}$. Recall that we know the dimension $p_n \equiv \mathfrak{D}(2,n)$ of the vector $\Psi_t^{(n)}$ from Eqn. (25), which also yields the dimension $\mathfrak{p}_n = p_1 + \cdots + p_n$ of the stacked vector $(\boldsymbol{\Psi}_t^{(1)}, \boldsymbol{\Psi}_t^{(2)}, \ldots, \boldsymbol{\Psi}_t^{(n)})^\top$. Consider the sequence of matrices $\{\boldsymbol{F}_n^{\mathfrak{p}_n\times\mathfrak{p}_n}\}_{n\in\mathbb{N}}$ given by

$$\boldsymbol{F}_n^{\mathfrak{p}_n\times\mathfrak{p}_n} = \begin{bmatrix} \boldsymbol{F}_{n-1}^{\mathfrak{p}_{n-1}\times\mathfrak{p}_{n-1}} & \boldsymbol{0}^{\mathfrak{p}_{n-1}\times p_n} \\ \boldsymbol{G}_n^{p_n\times\mathfrak{p}_{n-1}} & \boldsymbol{H}_n^{p_n\times p_n} \end{bmatrix}, \text{ with } \boldsymbol{H}_n^{p_n\times p_n} = \begin{bmatrix} \boldsymbol{M}^{(0,n)} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{K}^{(1,n-1)} & \boldsymbol{M}^{(1,n-1)} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{K}^{(n,0)} & \boldsymbol{M}^{(n,0)} \end{bmatrix}.$$

The matrices $\boldsymbol{G}_n^{p_n\times\mathfrak{p}_{n-1}}$ are not as elegantly expressed for general $n \in \mathbb{N}$, but can be explicitly obtained through Eqn. (21). We have used these matrices for $n = 1, 2$ to compute explicit moments in Appendix D.2, and in Appendix D.4 we give further details on these matrices for order $n = 3$. The stacked vector $(\boldsymbol{\Psi}_t^{(1)}, \boldsymbol{\Psi}_t^{(2)}, \ldots, \boldsymbol{\Psi}_t^{(n)})^\top$ satisfies the ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \boldsymbol{\Psi}_t^{(1)} \\ \vdots \\ \boldsymbol{\Psi}_t^{(n)} \end{bmatrix} = \boldsymbol{F}_n^{\mathfrak{p}_n\times\mathfrak{p}_n}\begin{bmatrix} \boldsymbol{\Psi}_t^{(1)} \\ \vdots \\ \boldsymbol{\Psi}_t^{(n)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{b}^{2\times 1} \\ \boldsymbol{0}^{(\mathfrak{p}_n - p_1)\times 1} \end{bmatrix}, \tag{43}$$

with initial condition given by $(\boldsymbol{\Psi}_0^{(1)}, \boldsymbol{\Psi}_0^{(2)}, \ldots, \boldsymbol{\Psi}_0^{(n)})^\top$. The solution is given in the following proposition, noting that taking the time derivative of Eqn. (44) immediately yields Eqn. (43).

Proposition 4. *If $\boldsymbol{H}_i^{p_i\times p_i}$ is invertible for all $i \in \{1,\ldots,n\}$, then*

$$\begin{bmatrix} \boldsymbol{\Psi}_t^{(1)} \\ \vdots \\ \boldsymbol{\Psi}_t^{(n)} \end{bmatrix} = e^{\boldsymbol{F}_n^{\mathfrak{p}_n\times\mathfrak{p}_n}t}\begin{bmatrix} \boldsymbol{\Psi}_0^{(1)} \\ \vdots \\ \boldsymbol{\Psi}_0^{(n)} \end{bmatrix} - \left(\boldsymbol{F}_n^{\mathfrak{p}_n\times\mathfrak{p}_n}\right)^{-1}\left(\boldsymbol{I}^{\mathfrak{p}_n\times\mathfrak{p}_n} - e^{\boldsymbol{F}_n^{\mathfrak{p}_n\times\mathfrak{p}_n}t}\right)\begin{bmatrix} \boldsymbol{b}^{2\times 1} \\ \boldsymbol{0}^{(\mathfrak{p}_n - p_1)\times 1} \end{bmatrix}. \tag{44}$$

A crucial computational advantage of Proposition 4 is that the evaluation of joint moments does not require integration of matrix exponentials as in Proposition 1. By considering the joint moments, the matrix in the associated ODE is more involved, but we reduced the non-homogeneous part of the ODE in Eqn. (43) to a constant, which allows for a closed-form solution. This in particular means that the run time of computing these transient moments does not increase in t, as opposed to the other computational methods. In the numerical computations of Section 7 we quantify the advantage in terms of computation time, compared to alternative approaches.

6. The Nearly Unstable Behavior

In this section, we analyze the stationary behavior of $\boldsymbol{\lambda}$ when the spectral radius of $\boldsymbol{H}$ approaches 1, see Eqn. (4), i.e., when the underlying process approaches criticality. This regime directly relates to the *heavy traffic* regime in queueing theory. In the setting considered, the dimension $d \in \mathbb{N}$ is general. For the sake of tractability, we impose the following 'symmetry assumption'.

Assumption 2. *For all $i \in [d]$,*

$$\alpha_i = \alpha \geqslant 0, \quad B_{1i} \stackrel{d}{=} \ldots \stackrel{d}{=} B_{di} \stackrel{d}{=} B_i, \quad \overline{\lambda}_i = \overline{\lambda} > 0, \tag{45}$$

where B_i are independent non-negative random variables with $\mathbb{E}[B_i^2] < \infty$.

This choice of parameters induces symmetry, since it implies that each component λ_i has the same base rate $\overline{\lambda}$, the same decay rate α, and that it is self- or cross-excited by all $B_1, \ldots, B_d$. The object of study is the Laplace transform

$$\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}) = \mathbb{E}\big[e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}}\big] = \mathbb{E}\Big[\prod_{i=1}^{d} e^{-s_i \lambda_i}\Big]. \tag{46}$$

The following result, proven in Appendix E, yields an explicit solution for $\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})$.

Lemma 1. *Assume Eqn.* (45) *and let $\beta_i(u) = \mathbb{E}[e^{-uB_i}]$ for any $u \geqslant 0$. Then, with $\overline{s} = s_1 + \cdots + s_d$, we have*

$$\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}) = \exp\Big(- \alpha\overline{\lambda} \int_0^{\overline{s}} \frac{u}{\alpha u + \sum_{i=1}^{d} \beta_i(u) - d} \mathrm{d}u\Big). \tag{47}$$

We now use the above lemma to derive the desired limit result in the nearly unstable case. Observe that the stability condition of the matrix $\boldsymbol{C}$, see Eqn. (4), having a maximum eigenvalue smaller than 1, is in our setting explicitly given by

$$\theta := \frac{1}{\alpha} \sum_{i=1}^{d} \mathbb{E}[B_i] < 1. \tag{48}$$

Furthermore, let $\sigma := 2\alpha\big(\sum_{i=1}^{d} \mathbb{E}[B_i^2]\big)^{-1}$. The following result is proven in Appendix E.

Theorem 3. *Under Assumption 2,*

$$\lim_{\theta \uparrow 1} \mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}(1-\theta)) = \Big(\frac{\sigma}{\sigma + \overline{s}}\Big)^{\sigma\overline{\lambda}}. \tag{49}$$

Theorem 3 shows that the limiting random vector has a (multivariate) Gamma distribution. It is immediately verified that, for any $i, j \in [d]$,

$$\lim_{\theta \uparrow 1} \mathrm{Cov}((1-\theta)\lambda_i, (1-\theta)\lambda_j) = \overline{\lambda}/\sigma, \tag{50}$$

by virtue of Eqn. (49). This yields the following result, writing $\Gamma(r, \lambda)$ for a Gamma distributed random variable with shape parameter $r > 0$ and scale parameter $\lambda > 0$. It follows immediately from Theorem 3 combined with Lévy's continuity theorem. The covariance expression follows from Eqn. (50).

Corollary 2. *Under Assumption 2, for a random vector $\boldsymbol{X}$, we have as $\theta \uparrow 1$*

$$(1-\theta)\boldsymbol{\lambda} \xrightarrow{d} \boldsymbol{X},$$

where each marginal $X_i \sim \Gamma(\sigma\overline{\lambda}, \sigma)$ and $\mathrm{Cov}(X_i, X_j) = \overline{\lambda}/\sigma$ for any $i, j \in [d]$.

7. Numerical Experiments

The objective of the results presented in the previous section is to be able to numerically evaluate moments. In this section we compare the resulting output, in terms of efficiency and accuracy, to two alternatives.

∘ The first alternative is based on *finite differences* (FD). The main idea is that we obtain approximations of moments by performing numerical differentiation of the relevant transform, using the characterizations in Theorem 1 and Corollary 1. As we have seen, the moments of interest can be obtained by appropriately differentiating the joint transform with respect to $\boldsymbol{s}$ and $\boldsymbol{z}$ and then setting $\boldsymbol{s} = \boldsymbol{0}$ and $\boldsymbol{z} = \boldsymbol{1}$. In the FD approach these derivatives are approximated by the corresponding (central) finite differences, parameterized the 'width parameter' $h > 0$. In our experiments, we assess how the precision of the approximations depends on h.

∘ The second alternative technique is based on *Monte Carlo simulation* (MC). To simulate the Hawkes process we use an algorithm based on Ogata's thinning algorithm; see [22] and [20, Algorithm 1.21] for details. The sampling mechanism is based on the cluster representation of [17, Definition 2]. Clearly, when relying on MC there is an evident tradeoff between precision and computational effort. Indeed, the well-known rule of thumb is that a reduction of the width of the confidence interval by a factor 2, requires the number of runs $m \in \mathbb{N}$ to be multiplied by a factor 4.

The first subsection focuses on computing moments of order $n = 1$ and $n = 2$ in the d-dimensional setting of Section 3.2. For $d = 3$ we compute these moments by evaluating the solutions of the ODEs derived in Section 3.2, which are used as benchmarks to compare the alternative methods to. In the second subsection, we consider a setting with $d = 2$, where we apply the results of the nested block-matrix developed in Section 5, allowing us to compute moments up to any order $n \in \mathbb{N}$. By this method, relying on analytical closed-form expressions, we compute moments of order up to three, which are later used as benchmarks.

The performance of the various approaches encompasses efficiency and precision, which we quantify in terms of *run time* and *error* (relative to the benchmark that we defined above, in the sequel abbreviated by BM), respectively. Two types of errors are distinguished, namely the Mean Absolute Error (MAE) and Mean Relative Error (MRE):

$$\mathrm{MAE} = \sum_{j=1}^{k} |m_j^{(\mathrm{BM})} - m_j^{(\mathrm{FD/MC})}|, \quad \mathrm{MRE} = \sum_{j=1}^{k} \frac{|m_j^{(\mathrm{BM})} - m_j^{(\mathrm{FD/MC})}|}{m_j^{(\mathrm{BM})}}, \tag{51}$$

where m_j denotes our BM value of the j-th moment, k is the number of moments computed, and the superscript indicates the computational method used. The general conclusion of this section is that the experiments systematically reveal that the benchmarks outperform the alternative approaches, in terms of efficiency and accuracy, typically by a large margin.

7.1. **Multivariate.** We consider an example of dimension $d = 3$, in which we numerically evaluate the first and second order moments of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$. We start by taking $t = 5$, to later study the impact of the value of t. The marks are exponentially distributed: for any combination $i, j \in \{1, 2, 3\}$, we set $B_{ij} \sim \mathrm{Exp}(b_{ij})$ for some $b_{ij} > 0$, which are (for simplicity) assumed independent. In the experiments we take

$$\overline{\boldsymbol{\lambda}} = \begin{bmatrix} 0.3 \\ 1 \\ 0.5 \end{bmatrix}, \quad \mathbb{E}\boldsymbol{B} = \begin{bmatrix} 0.5 & 0.3 & 0.4 \\ 0.7 & 0.5 & 0.5 \\ 0.4 & 0.2 & 0.5 \end{bmatrix}, \quad \boldsymbol{D}_\alpha = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 2.5 \end{bmatrix}, \quad \boldsymbol{D}_\mu = \begin{bmatrix} 1.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

One readily verifies that for these parameters the stability condition of Assumption 1 is met.

Recall that the stacked vector $\boldsymbol{\Sigma}_t^{(1)}$ and the stacked matrix $\boldsymbol{\Sigma}_t^{(2)}$ contain all the first order moments and combinations of second order moments, respectively; see Eqns. (64) and (67). In this subsection, the benchmark BM corresponds to the solution of the vector- and matrix-valued ODEs as given in Eqns. (65) and (68), obtained using the `SciPy` package in `Python`. We used the default precision of the `SciPy` ODE solver; the output presented in the next subsection indicates that this provides sufficiently precise results.

Table 1, displaying the resulting run times and errors, quantifies the superior performance of our approach. In each of the settings, the run time RT is reliably estimated by performing the experiment sufficiently often and taking the average of the corresponding run times. The table shows that the benchmark ODE method is faster than the FD method, and orders of magnitude faster than the Monte Carlo simulation, where the latter method in addition typically yields estimates with substantial errors (where the number of simulation runs is $m = 10^3$). We see that in the FD method smaller values of h lead to lower run times: in this method, we vary the arguments $\boldsymbol{s}$ and $\boldsymbol{z}$ with h when evaluating the joint transform, which is faster for smaller values of h. Furthermore, observe that the MAE and MRE are not monotone in h: for larger h the derivative is poorly approximated by the finite difference, while for smaller h numerical stability issues have a detrimental effect. There is an optimal width where the error is smallest, which in our instance happens to be around $h = 10^{-3}$.

To assess whether the effects observed in the previous experiment hold in general, we have performed experiments with a set of intrinsically different parameter settings. In these experiments, we study run times and errors, while we fix the 'width parameter' at $h = 10^{-3}$

	BM	FD				MC			
n	RT	h	RT	MAE	MRE	m	RT	MAE	MRE
1	$7.17 \cdot 10^{-3}$	10^{-2}	$5.26 \cdot 10^{-2}$	$1.27 \cdot 10^{-3}$	$3.42 \cdot 10^{-4}$	10^2	9	$8.92 \cdot 10^{-1}$	$4.98 \cdot 10^{-1}$
	·	10^{-3}	$3.94 \cdot 10^{-2}$	$1.81 \cdot 10^{-4}$	$9.84 \cdot 10^{-5}$	10^3	96	$2.41 \cdot 10^{-1}$	$1.43 \cdot 10^{-1}$
	·	10^{-4}	$3.35 \cdot 10^{-2}$	$6.60 \cdot 10^{-4}$	$3.69 \cdot 10^{-4}$	10^4	956	$6.87 \cdot 10^{-2}$	$3.68 \cdot 10^{-2}$
2	$9.65 \cdot 10^{-2}$	10^{-2}	$3.29 \cdot 10^{-1}$	$9.96 \cdot 10^{-2}$	$1.17 \cdot 10^{-2}$	10^2	13	$2.48 \cdot 10^{1}$	$4.56 \cdot 10^{0}$
	·	10^{-3}	$2.62 \cdot 10^{-1}$	$2.03 \cdot 10^{-3}$	$4.18 \cdot 10^{-4}$	10^3	115	$3.89 \cdot 10^{0}$	$7.69 \cdot 10^{-1}$
	·	10^{-4}	$2.20 \cdot 10^{-1}$	$1.13 \cdot 10^{-2}$	$1.18 \cdot 10^{-3}$	10^4	1025	$9.07 \cdot 10^{-1}$	$2.42 \cdot 10^{-1}$

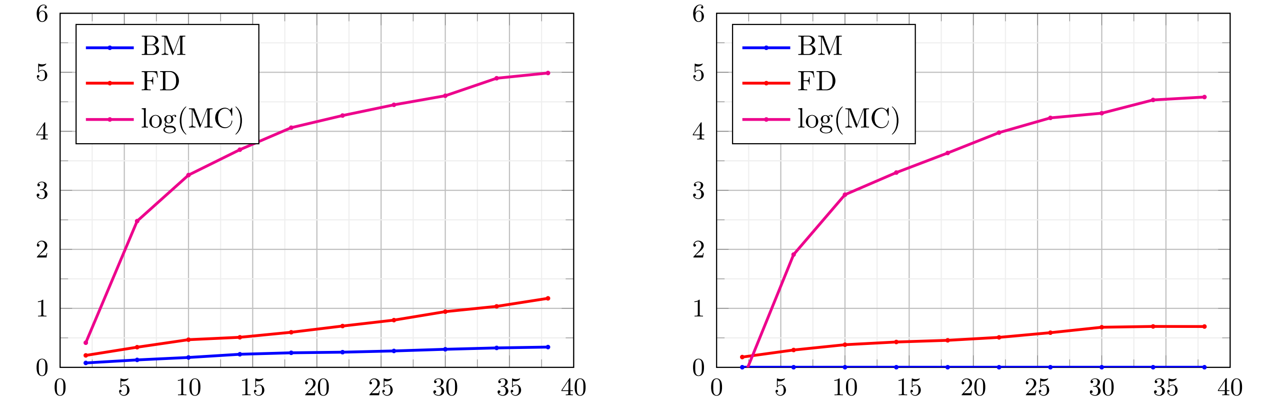
Table 1. *Run times (RT) in seconds and errors (MAE, MRE) for first ($n = 1$) and second ($n = 2$) order moments: performance of the benchmark ODE-based method relative to FD and MC.*

and number of simulation runs at $m = 10^3$. Since varying each entry in each of the vectors and matrices would lead to a large set of instances, we decided to focus on altering only the parameters directly pertaining to $\lambda_1(\cdot)$ and $Q_1(\cdot)$, while respecting Assumption 1; note that the effect will propagate to other components due to cross-excitation. Table 2 shows the resulting run times and errors. The main conclusion is that the experiments reveal that, uniformly across all instances, the benchmark ODE-based method remains the fastest, with a run time that is hardly affected by the parameters chosen. We note that increasing the value of $\mathbb{E}[B_{11}]$ or decreasing the value of α_1 results in the system approaching the boundary of the stability condition in Assumption 1, thus leading to larger relative errors.

	BM	FD			MC		
Parameter	RT	RT	MAE	MRE	RT	MAE	MRE
$\overline{\lambda}_1 = 3$	$1.13 \cdot 10^{-1}$	$3.04 \cdot 10^{-1}$	$2.48 \cdot 10^{-2}$	$7.44 \cdot 10^{-4}$	286	$7.19 \cdot 10^{0}$	$3.34 \cdot 10^{-1}$
$\overline{\lambda}_1 = 5$	$1.20 \cdot 10^{-1}$	$3.10 \cdot 10^{-1}$	$7.68 \cdot 10^{-2}$	$1.02 \cdot 10^{-3}$	432	$1.47 \cdot 10^{1}$	$3.98 \cdot 10^{-1}$
$\overline{\lambda}_1 = 10$	$1.36 \cdot 10^{-1}$	$3.32 \cdot 10^{-1}$	$5.37 \cdot 10^{-1}$	$2.09 \cdot 10^{-3}$	812	$1.18 \cdot 10^{2}$	$6.19 \cdot 10^{-1}$
$\mathbb{E}B_{11} = 1$	$1.04 \cdot 10^{-1}$	$3.05 \cdot 10^{-1}$	$1.02 \cdot 10^{-2}$	$1.14 \cdot 10^{-3}$	118	$5.48 \cdot 10^{0}$	$7.55 \cdot 10^{-1}$
$\mathbb{E}B_{11} = 1.3$	$1.06 \cdot 10^{-1}$	$3.07 \cdot 10^{-1}$	$1.52 \cdot 10^{-2}$	$1.07 \cdot 10^{-3}$	131	$1.12 \cdot 10^{1}$	$9.38 \cdot 10^{-1}$
$\mathbb{E}B_{11} = 1.6$	$1.02 \cdot 10^{-1}$	$3.08 \cdot 10^{-1}$	$1.19 \cdot 10^{-1}$	$2.62 \cdot 10^{-3}$	152	$8.72 \cdot 10^{1}$	$2.38 \cdot 10^{0}$
$\alpha_1 = 1$	$1.03 \cdot 10^{-1}$	$2.97 \cdot 10^{-1}$	$7.91 \cdot 10^{-3}$	$5.11 \cdot 10^{-3}$	130	$1.29 \cdot 10^{1}$	$1.09 \cdot 10^{0}$
$\alpha_1 = 3$	$1.11 \cdot 10^{-1}$	$3.27 \cdot 10^{-1}$	$2.89 \cdot 10^{-3}$	$1.92 \cdot 10^{-3}$	93	$5.85 \cdot 10^{0}$	$2.04 \cdot 10^{0}$
$\alpha_1 = 10$	$1.73 \cdot 10^{-1}$	$5.45 \cdot 10^{-1}$	$1.14 \cdot 10^{-2}$	$8.89 \cdot 10^{-3}$	80	$2.29 \cdot 10^{0}$	$9.83 \cdot 10^{-1}$
$\mu_1 = 0.5$	$1.01 \cdot 10^{-1}$	$3.01 \cdot 10^{-1}$	$2.49 \cdot 10^{-3}$	$5.05 \cdot 10^{-4}$	98	$3.84 \cdot 10^{0}$	$6.30 \cdot 10^{-1}$
$\mu_1 = 2$	$1.06 \cdot 10^{-1}$	$3.04 \cdot 10^{-1}$	$2.23 \cdot 10^{-3}$	$5.82 \cdot 10^{-4}$	97	$4.98 \cdot 10^{0}$	$1.25 \cdot 10^{0}$
$\mu_1 = 5$	$1.26 \cdot 10^{-1}$	$3.18 \cdot 10^{-1}$	$2.42 \cdot 10^{-3}$	$1.12 \cdot 10^{-3}$	99	$5.22 \cdot 10^{0}$	$1.45 \cdot 10^{0}$

Table 2. *Run times (RT) in seconds and errors (MAE, MRE) for combined first and second order moments in the trivariate setting: effect of parameter changes of the benchmark ODE-based method relative to FD and MC.*

We also studied the effect of varying the time parameter t on the run times. Recall that the FD method uses the (conditional) joint transform, where the latter requires solving systems of ODEs. Figure 1b shows that the run times of the ODE-based method and the FD method scale effectively linearly with t, with the ODE method having the smaller slope. The run time for MC increases superlinearly; we took its logarithm to be able to show it in the same plot. This superlinear behavior is an inherent consequence of the branching structure underlying the Hawkes process (relied upon in Ogata's algorithm). We note that this qualitative behavior is observed for all choices of parameters that we considered.



(A) Trivariate setting: up to 2nd-order moments (B) Bivariate setting: up to 3rd-order moments

Figure 1. Run times of the BM, FD, and MC methods for computing moments. (a) Trivariate setting (up to second order). (b) Bivariate setting (up to third order).

7.2. **Bivariate.** In this subsection, we compute for the bivariate setting ($d = 2$) the transient moments of $\boldsymbol{Q}(t) = (Q_1(t), Q_2(t))$ and $\boldsymbol{\lambda}(t) = (\lambda_1(t), \lambda_2(t))$, of orders $n = 1, 2, 3$. Again we take $t = 5$, but later assess the effect of the choice of t. As before the random marks are exponentially distributed, i.e., for $i, j \in \{1, 2\}$, we set $B_{ij} \sim \text{Exp}(b_{ij})$ for some $b_{ij} > 0$, with independence between the B_{ij}. The parameters are

$$\overline{\boldsymbol{\lambda}} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, \quad \mathbb{E}\boldsymbol{B} = \begin{bmatrix} 1.5 & 0.5 \\ 0.75 & 1.25 \end{bmatrix}, \quad \boldsymbol{D}_\alpha = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}, \quad \boldsymbol{D}_\mu = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

It can be verified that the stability condition of Eqn. (4), which in this bivariate case reads $(\alpha_1 - \mathbb{E}[B_{11}])(\alpha_2 - \mathbb{E}[B_{22}]) > \mathbb{E}[B_{12}]\mathbb{E}[B_{21}]$, is met.

We compute all the first, second and third order moments, i.e., all entries of the stacked vectors $\boldsymbol{\Psi}_t^{(1)}$, $\boldsymbol{\Psi}_t^{(2)}$, and $\boldsymbol{\Psi}_t^{(3)}$. We use the main result of Section 5, namely Proposition 4, as our benchmark. This result exploits the block-matrix structure, by which we can simultaneously compute moments of multiple orders, thus greatly increasing the computational performance. As it turns out, the output hardly differs from what is obtained by the ODE-based approach of the previous subsection (i.e., a difference in the order of 10^{-8}). The difference in computational effort, however, is substantial: for this instance the block-matrix method is about 200 times faster.

Table 3 shows that the BM method is much faster than FD and MC, especially for second and third order moments (with again the number of simulation runs being $m = 10^3$). We also see that the absolute and relative errors of FD and MC significantly grow as the order of moments increase. Particularly for the third order moments, the poor stability of the FD method significantly degrades the performance, as can be seen by the variability of the error when changing the precision parameter h.

	BM	FD				MC			
n	RT	h	RT	MAE	MRE	m	RT	MAE	MRE
1	$4.77 \cdot 10^{-4}$	10^{-2}	$4.42 \cdot 10^{-2}$	$2.55 \cdot 10^{-3}$	$1.12 \cdot 10^{-3}$	10^{2}	5	$7.16 \cdot 10^{-1}$	$3.32 \cdot 10^{-1}$
	·	10^{-3}	$3.31 \cdot 10^{-2}$	$8.55 \cdot 10^{-5}$	$4.61 \cdot 10^{-5}$	10^{3}	62	$3.19 \cdot 10^{-1}$	$1.49 \cdot 10^{-1}$
	·	10^{-4}	$2.97 \cdot 10^{-2}$	$4.38 \cdot 10^{-4}$	$2.64 \cdot 10^{-4}$	10^{4}	589	$1.69 \cdot 10^{-2}$	$7.30 \cdot 10^{-3}$
2	$5.61 \cdot 10^{-4}$	10^{-2}	$8.12 \cdot 10^{-2}$	$2.70 \cdot 10^{-1}$	$2.45 \cdot 10^{-2}$	10^{2}	6	$2.98 \cdot 10^{1}$	$2.76 \cdot 10^{0}$
	·	10^{-3}	$6.42 \cdot 10^{-2}$	$2.03 \cdot 10^{-3}$	$4.18 \cdot 10^{-4}$	10^{3}	67	$8.01 \cdot 10^{0}$	$9.71 \cdot 10^{-1}$
	·	10^{-4}	$5.73 \cdot 10^{-2}$	$1.14 \cdot 10^{-2}$	$1.38 \cdot 10^{-3}$	10^{4}	631	$5.36 \cdot 10^{0}$	$6.92 \cdot 10^{-1}$
3	$9.26 \cdot 10^{-4}$	10^{-2}	$2.49 \cdot 10^{-1}$	$2.23 \cdot 10^{2}$	$1.77 \cdot 10^{0}$	10^{2}	7	$9.57 \cdot 10^{2}$	$1.09 \cdot 10^{1}$
	·	10^{-3}	$2.07 \cdot 10^{-1}$	$2.99 \cdot 10^{2}$	$5.61 \cdot 10^{1}$	10^{3}	69	$2.21 \cdot 10^{2}$	$2.87 \cdot 10^{0}$
	·	10^{-4}	$1.72 \cdot 10^{-1}$	$1.34 \cdot 10^{6}$	$3.34 \cdot 10^{4}$	10^{4}	639	$3.98 \cdot 10^{1}$	$4.47 \cdot 10^{-1}$

Table 3. *Run times (RT) in seconds and errors (MAE, MRE) for first* ($n = 1$), *second* ($n = 2$), *and third* ($n = 3$) *order moments in the bivariate setting: comparison of the benchmark block-matrix method relative to FD and MC.*

Figure 1a shows the effect of the time parameter t. As before, the run time of FD scales linearly with t, and that of MC superlinearly. This should be contrasted with the attractive feature of the block-matrix method that its run time does not depend on t.

We now consider the block-matrix stationary moments. The first and second order stationary moments can be immediately obtained from the results of Appendix C.2, by solving the associated Sylvester matrix equations. Obvious alternatives when not knowing these stationary moments, amount to picking a 'large' value of t in the FD and MC methods. These alternative methods have two intrinsic drawbacks: (1) run times increase in t, and (2) we do not know *a priori* what value of t guarantees that the error made is sufficiently small. As we already saw that MC is typically outperformed by FD, we focus on FD only. A pragmatic way to select a sufficiently large value of t, is to compute the FD-based approximation of first and second order transient moments for successive *integer* values of t, until the difference of the respective MREs is smaller than some given threshold precision level ϵ. We compare the resulting approximation to our benchmark, i.e., the values obtained by solving the Sylvester matrix equations, so as to quantify the error made.

Table 4 presents the results for the FD method with precision level $\epsilon = 0.01$. We have performed the above procedure for different choices of parameters, where the parameter that is altered is given in the table. Note that the benchmark method is exact and provides near-instant response. Observe that for specific sets of parameters there is a substantial effect on the value of t (i.e., the value of the time parameter at which the procedure terminates), the run time, and the MRE, in particular when the parameters are close to the boundary of the stability condition in Assumption 1, e.g., $\mathbb{E}[B_{11}] = 2.25$ or $\alpha_1 = 2.1$.

In Appendix F we present a numerical evaluation of the objects featured in Section 3.1.

8. Concluding Remarks

This paper has studied multivariate Hawkes-fed Markovian infinite-server queues, which can be alternatively interpreted as population processes. Our objective was to devise accurate and

	FD		
Parameter	t	RT	MRE
$\overline{\lambda}_1 = 3$	18	$3.25 \cdot 10^0$	$2.17 \cdot 10^{-2}$
$\overline{\lambda}_1 = 10$	18	$3.38 \cdot 10^0$	$1.94 \cdot 10^{-2}$
$\mathbb{E}B_{11} = 0.5$	14	$2.75 \cdot 10^0$	$1.51 \cdot 10^{-2}$
$\mathbb{E}B_{11} = 2.25$	43	$1.08 \cdot 10^1$	$6.87 \cdot 10^{-2}$
$\alpha_1 = 2.1$	86	$2.89 \cdot 10^1$	$1.58 \cdot 10^{-1}$
$\alpha_1 = 5$	15	$3.77 \cdot 10^0$	$1.55 \cdot 10^{-2}$
$\mu_1 = 0.5$	19	$3.97 \cdot 10^0$	$2.11 \cdot 10^{-2}$
$\mu_1 = 5$	18	$3.38 \cdot 10^0$	$2.84 \cdot 10^{-2}$

Table 4. *Run times (RT) in seconds until first and second order transient moments approximate stationary moments in the bivariate setting: FD method with precision level* $\epsilon = 0.01$.

efficient algorithms to compute transient and stationary moments. We succeeded in doing so, heavily relying on having access to the joint transform of the Hawkes intensity process and the population process. When the multivariate Hawkes process is of general dimension d, this transform is expressed in terms of systems of ODEs, allowing for the computation of joint moments. This includes joint moments where the components pertain to the same as well as to different points in time, thus also covering the evaluation of the processes' autocovariance functions. We then proceeded by deriving expressions for the first and second order, transient and stationary moments for the d-dimensional processes. Next, in the 2-dimensional setting we derived a recursive procedure, revealing a block-matrix structure for the computation of moments of any order. Our numerical experiments show that our approach outperforms existing alternatives: it produces highly accurate results with computation times that are virtually negligible.

This paper considering Markovian multivariate Hawkes processes and associated population processes, we conclude this section by a brief discussion of potential extensions to more general classes of processes. In the first place, where in this work we exclusively focus on exponential decay functions, one could work with more general non-increasing and integrable decay functions. When doing so, one leaves the Markovian setting, rendering the results of this paper not applicable. Instead, the process may be analyzed through the cluster representation, first described in [16], where the Hawkes process is described as a Poisson cluster process. This approach has been followed in [17] to study distributional properties in a multivariate setting, including the multivariate Hawkes process as well as the associated population process. Another extension concerns the nonlinear case of e.g. [3, 25], entailing that the Hawkes process is not even a Poisson cluster process and therefore requires entirely different analysis techniques.

References

[1] Aït-Sahalia, Y., Cacho-Diaz, J. A. & Laeven, R. J. A. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics* 117, 585-606.

[2] Aït-Sahalia, Y., Laeven, R. J. A. & Pelizzon, L. (2014). Mutual excitation in Eurozone sovereign CDS. *Journal of Econometrics* 183, 151-167.

[3] Brémaud, P. & Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Annals of Probability* 24, 1562-1588.

[4] Chen, X. & Hong, G. (2023). Steady-state analysis and online learning for queues with Hawkes arrivals. Preprint. Available at `https://arxiv.org/pdf/2311.02577.pdf`.

[5] Chen, Z., Dassios, A., Kuan, V., Lim, J. W., Qu, Y., Surya, B. & Zhao, H. (2020). A two-phase dynamic contagion model for COVID-19. Preprint. Available at `https://www.lse.ac.uk/Statistics/Assets/Documents/Paper-Covid-19-DCP.pdf`.

[6] Cui, L., Hawkes, A. & Yi, H. (2020). An elementary derivation of moments of Hawkes processes. *Advances in Applied Probability* 52, 102-137.

[7] Da Fonseca, J. & Zaatour, R. (2015). Clustering and mean reversion in a Hawkes microstructure model. *Journal of Futures Markets* 35, 813-838.

[8] Daley, D. J. & Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Volume 1: Elementary Theory and Methods.* Second Edition. Springer Science and Business Media, New York.

[9] Dassios, A. & Zhou, H. (2011). A dynamic contagion process. *Advances in Applied Probability* 43, 814-846.

[10] Daw, A., Castellanos, A., Yom-Tov, G., Pender, J. & Gruendlinger, L. (2025). The co-production of service: Modeling service times in contact centers using Hawkes processes. *Management Science* 71, 1865-1888.

[11] Daw, A. & Pender, J. (2023). Matrix calculations for moments of Markov processes. *Advances in Applied Probability* 55, 126-150.

[12] Daw, A. & Pender, J. (2018). Queues driven by Hawkes processes. *Stochastic Systems* 8, 192-229.

[13] Du, D. & Luo, D. (2019). The pricing of jump propagation: Evidence from spot and options markets. *Management Science* 65, 2360-2387.

[14] Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society.* Series B (Methodological), 438-443.

[15] Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 83-90.

[16] Hawkes, A. G. & Oakes, D. (1974). A cluster representation of a self-exciting process. *Journal of Applied Probability* 11, 493-503.

[17] Karim, R. S., Laeven, R. J. A. & Mandjes, M. (2021). Exact and asymptotic analysis of general multivariate Hawkes processes and induced population processes. Preprint. Available at `https://arxiv.org/pdf/2106.03560.pdf`.

[18] Koops, D. T., Saxena, M., Boxma, O. J. & Mandjes, M. (2018). Infinite-server queues with Hawkes input. *Journal of Applied Probability* 55, 920-943.

[19] Laub, P. J., Taimre, T. & Pollett, P. K. (2015). Hawkes Processes. Preprint. Available at `https://arxiv.org/abs/1507.02822`.

[20] Liniger, T. J. (2009). *Multivariate Hawkes Processes.* PhD thesis, ETH Zürich.

[21] Oakes, D. (1975). The Markovian self-exciting process. *Journal of Applied Probability* 12, 69-77.

[22] Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory* 27, 23-31.

[23] Rizoiu, M.-A., Mishra, S. Kong, Q., Carman, M. & Xie, L. (2018). SIR-Hawkes: Linking epidemic models and Hawkes processes to model diffusions in finite populations. Preprint. Available at `https://arxiv.org/abs/1711.01679`

[24] Xu, L., Duan, J. & Whinston, A. (2014). Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Science* 60, 1392-1412.

[25] Zhu, L. (2013). Central limit theorem for nonlinear Hawkes processes. *Journal of Applied Probability* 50, 760-771.

Appendix A. Proofs of Theorems 1 and 2

Proof of Theorem 1. The proof is comprised of a number of steps. First, we use the Markov property on the distribution function of the joint process $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$, next we take partial derivatives to obtain an expression for the density. Then, we consecutively apply the Laplace and $\boldsymbol{z}$-transform to obtain a PDE. Finally, we use the method of characteristics to obtain a system of ODEs. We describe below the main steps; some lengthy technical computations have been put into Appendix B.

We note that the probabilities considered in this proof are conditional on the value of the processes at time t_0, i.e. $\boldsymbol{Q}(t_0) = \boldsymbol{Q}_0$ and $\boldsymbol{\lambda}(t_0) = \boldsymbol{\lambda}_0$. To start off, for $t \in \mathbb{R}_+$, $\boldsymbol{k} \in \mathbb{N}_+^d$ and $\boldsymbol{\nu} \in \mathbb{R}_+^d$, set

$$F(t, \boldsymbol{\nu}, \boldsymbol{k}) = \mathbb{P}(\boldsymbol{\lambda}(t) \leqslant \boldsymbol{\nu}, \boldsymbol{Q}(t) = \boldsymbol{k}), \quad \frac{\partial F(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \boldsymbol{\nu}} = \left(\frac{\partial F(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_1}, \cdots, \frac{\partial F(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_d} \right)^\top. \tag{52}$$

Also define

$$f(t, \boldsymbol{\nu}, \boldsymbol{k}) = \frac{\partial^d F(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_1 \cdots \partial \nu_d}, \tag{53}$$

as the joint density of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$. For some $\delta > 0$, consider the probability

$$F(t + \delta, \boldsymbol{\nu} - \boldsymbol{\alpha} \odot (\boldsymbol{\nu} - \overline{\boldsymbol{\lambda}})\delta, \boldsymbol{k}), \tag{54}$$

where the interpretation of the term $\boldsymbol{\nu} - \boldsymbol{\alpha} \odot (\boldsymbol{\nu} - \overline{\boldsymbol{\lambda}})$ is a decay factor, in the sense that for small δ, no new arrival of a point in $(t, t+\delta]$ makes the intensity $\boldsymbol{\lambda}(\cdot)$ decay with rate $\boldsymbol{\alpha}$ back to the mean reversion level $\overline{\boldsymbol{\lambda}}$. To compute this probability, we apply the Markov property to Eqn. (54), which leaves us to consider the possibilities to get in the state of exactly $\boldsymbol{k}$ active points and intensity equal to $\boldsymbol{\nu} - \boldsymbol{\alpha} \odot (\boldsymbol{\nu} - \overline{\boldsymbol{\lambda}})$. There are three distinct ways to get to this state at time $t+\delta$ from time t: we have exactly $\boldsymbol{k}$ active points with no arrivals or departures; we have $\boldsymbol{k} - \boldsymbol{e}_j$ active points and exactly one arrival in component j; or we have $\boldsymbol{k} + \boldsymbol{e}_j$ active points and one departure in component j. This yields, up to $o(\delta)$ terms, that

$$\begin{aligned}
&F(t + \delta, \boldsymbol{\nu} - \boldsymbol{\alpha} \odot (\boldsymbol{\nu} - \overline{\boldsymbol{\lambda}})\delta, \boldsymbol{k}) \\
&= \sum_{j=1}^d \int_0^{\nu_d} \cdots \int_0^{\nu_1} \delta y_j f(t, \boldsymbol{y}, \boldsymbol{k} - \boldsymbol{e}_j) \mathbb{P}(\boldsymbol{B}_j \leqslant \boldsymbol{\nu} - \boldsymbol{y}) \mathrm{d}y_1 \cdots \mathrm{d}y_d \\
&\quad + \sum_{j=1}^d (k_j + 1) \delta \mu_j F(t, \boldsymbol{\nu}, \boldsymbol{k} + \boldsymbol{e}_j) \\
&\quad + \int_0^{\nu_d} \cdots \int_0^{\nu_1} (1 - \sum_{j=1}^d \delta \mu_j k_j - \sum_{j=1}^d \delta y_j) f(t, \boldsymbol{y}, \boldsymbol{k}) \mathrm{d}y_1 \cdots \mathrm{d}y_d + o(\delta).
\end{aligned}$$

Subtracting $F(t, \boldsymbol{k}, \boldsymbol{\lambda})$ on both sides, dividing by δ, and taking $\delta \downarrow 0$ yields

$$\begin{aligned}
&\frac{\partial F(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial t} - \big(\boldsymbol{\alpha} \odot (\boldsymbol{\nu} - \overline{\boldsymbol{\lambda}})\big)^\top \frac{\partial F(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \boldsymbol{\nu}} \\
&= \sum_{j=1}^d \int_0^{\nu_d} \cdots \int_0^{\nu_1} y_j f(t, \boldsymbol{y}, \boldsymbol{k} - \boldsymbol{e}_j) \mathbb{P}(\boldsymbol{B}_j \leqslant \boldsymbol{\nu} - \boldsymbol{y}) \mathrm{d}y_1 \cdots \mathrm{d}y_d \\
&\quad + \sum_{j=1}^d (k_j + 1) \mu_j F(t, \boldsymbol{\nu}, \boldsymbol{k} + \boldsymbol{e}_j) - \sum_{j=1}^d \int_0^{\nu_d} \cdots \int_0^{\nu_1} (\mu_j k_j + y_j) f(t, \boldsymbol{y}, \boldsymbol{k}) \mathrm{d}y_1 \cdots \mathrm{d}y_d,
\end{aligned}$$

where the left-hand side follows from the definition of the directional derivative. Next, we take the partial derivatives with respect to $\nu_1, \ldots, \nu_d$, so as to rewrite above equation in terms of the probability density function $f(t, \boldsymbol{\nu}, \boldsymbol{k})$. By the definitions of F and f given in Eqns. (52) and (53), and using Leibniz' integral rule on the integral terms, we obtain

$$\begin{aligned}
&\frac{\partial f(t,\boldsymbol{\nu},\boldsymbol{k})}{\partial t} - \sum_{j=1}^{d} \alpha_j \frac{\partial}{\partial \nu_j} \nu_j f(t,\boldsymbol{\nu},\boldsymbol{k}) + \sum_{j=1}^{d} \alpha_j \overline{\lambda}_j \frac{\partial f(t,\boldsymbol{\nu},\boldsymbol{k})}{\partial \nu_j} \\
&= \sum_{j=1}^{d} \int_0^{\nu_d} \cdots \int_0^{\nu_1} y_j f(t,\boldsymbol{y},\boldsymbol{k}-\boldsymbol{e}_j) \frac{\partial^d}{\partial \nu_1 \cdots \partial \nu_d} \mathbb{P}(\boldsymbol{B}_j \leqslant \boldsymbol{\nu} - \boldsymbol{y}) \mathrm{d}y_1 \cdots \mathrm{d}y_d \\
&\quad + \sum_{j=1}^{d} \Big(f(t,\boldsymbol{\nu},\boldsymbol{k}+\boldsymbol{e}_j)(k_j+1)\mu_j - f(t,\boldsymbol{\nu},\boldsymbol{k})(k_j\mu_j + \nu_j) \Big).
\end{aligned} \tag{55}$$

Denote the d-dimensional Laplace transform with respect to $\boldsymbol{\nu}$ by

$$\xi(t,\boldsymbol{s},\boldsymbol{k}) := \mathcal{L}(f(t,\boldsymbol{\nu},\boldsymbol{k}))(\boldsymbol{s}) = \int_0^\infty \cdots \int_0^\infty e^{-\boldsymbol{s}^\top \boldsymbol{\nu}} f(t,\boldsymbol{\nu},\boldsymbol{k}) \mathrm{d}\nu_1 \cdots \mathrm{d}\nu_d.$$

Taking the Laplace transform of Eqn. (55) yields

$$\begin{aligned}
&\frac{\partial \xi(t,\boldsymbol{s},\boldsymbol{k})}{\partial t} + \sum_{j=1}^{d} \alpha_j s_j \frac{\partial \xi(t,\boldsymbol{s},\boldsymbol{k})}{\partial s_j} + \sum_{j=1}^{d} \alpha_j \overline{\lambda}_j s_j \xi(t,\boldsymbol{s},\boldsymbol{k}) \qquad (56)\\
&= \sum_{j=1}^{d} \Big(-\frac{\partial \xi(t,\boldsymbol{s},\boldsymbol{k}-\boldsymbol{e}_j)}{\partial s_j} \beta_j(\boldsymbol{s}) + (k_j+1)\mu_j \xi(t,\boldsymbol{s},\boldsymbol{k}+\boldsymbol{e}_j) - k_j\mu_j \xi(t,\boldsymbol{s},\boldsymbol{k}) + \frac{\partial \xi(t,\boldsymbol{s},\boldsymbol{k})}{\partial s_j} \Big);
\end{aligned}$$

see Appendix B.1 for the term-by-term derivation. The computations boil down to applying integration by parts, convolution arguments and the properties of F and f. Rewriting the expression in such a way that all derivatives end up on one side,

$$\begin{aligned}
&\frac{\partial \xi(t,\boldsymbol{s},\boldsymbol{k})}{\partial t} + \sum_{j=1}^{d} (\alpha_j s_j - 1) \frac{\partial \xi(t,\boldsymbol{s},\boldsymbol{k})}{\partial s_j} + \sum_{j=1}^{d} \frac{\partial \xi(t,\boldsymbol{s},\boldsymbol{k}-\boldsymbol{e}_j)}{\partial s_j} \beta_j(\boldsymbol{s}) \\
&= \sum_{j=1}^{d} \Big((k_j+1)\mu_j \xi(t,\boldsymbol{s},\boldsymbol{k}+\boldsymbol{e}_j) - k_j\mu_j \xi(t,\boldsymbol{s},\boldsymbol{k}) - \alpha_j \overline{\lambda}_j s_j \xi(t,\boldsymbol{s},\boldsymbol{k}) \Big).
\end{aligned} \tag{57}$$

Next, we rewrite equation (57) by taking the $\boldsymbol{z}$-transform, which gives us the joint transform introduced in Eqn. (6), i.e.,

$$\zeta_{t_0}(t,\boldsymbol{s},\boldsymbol{z}) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_d=0}^{\infty} z_1^{k_1} \cdots z_d^{k_d} \xi(t,\boldsymbol{s},\boldsymbol{k}) = \mathbb{E}_{t_0}\Big[e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{i=1}^{d} z_i^{Q_i(t)} \Big],$$

yielding

$$\begin{aligned}
&\frac{\partial \zeta_{t_0}(t,\boldsymbol{s},\boldsymbol{z})}{\partial t} + \sum_{j=1}^{d} \big(\alpha_j s_j + z_j \beta_j(\boldsymbol{s}) - 1 \big) \frac{\partial \zeta_{t_0}(t,\boldsymbol{s},\boldsymbol{z})}{\partial s_j} + \sum_{j=1}^{d} \big(\mu_j (z_j - 1) \big) \frac{\partial \zeta_{t_0}(t,\boldsymbol{s},\boldsymbol{z})}{\partial z_j} \\
&\qquad = -\zeta_{t_0}(t,\boldsymbol{s},\boldsymbol{z}) \sum_{j=1}^{d} \alpha_j \overline{\lambda}_j s_j,
\end{aligned} \tag{58}$$

where we added the subscript t_0 to emphasize the dependence on this initial time value. We refer to Appendix B.1 for the term-by-term derivation.

By employing the method of characteristics, we can rewrite the PDE in Eqn. (58) into a system of ODEs. To that end, consider a curve in $\mathbb{R}^{2d}$ parameterized by $(\widehat{\boldsymbol{s}}(u), \widehat{\boldsymbol{z}}(u))$ as a function of u, where $t_0 \leqslant u \leqslant t$, which terminates at the set of parameters $(\boldsymbol{s}, \boldsymbol{z})$, i.e. $(\widehat{\boldsymbol{s}}(t), \widehat{\boldsymbol{z}}(t)) = (\boldsymbol{s}, \boldsymbol{z})$. Since we have a first-order PDE, we easily obtain the characteristic system of ODEs by

$$\begin{aligned} \frac{\mathrm{d}\widehat{s}_j(u)}{\mathrm{d}u} &= \alpha_j \widehat{s}_j(u) + \widehat{z}_j(u)\beta_j(\widehat{\boldsymbol{s}}(u)) - 1, \\ \frac{\mathrm{d}\widehat{z}_j(u)}{\mathrm{d}u} &= \mu_j(\widehat{z}_j(u) - 1), \end{aligned} \tag{59}$$

for each $j \in [d]$. For $\widehat{z}_j(\cdot)$, the solution can be directly computed as

$$\widehat{z}_j(u) = 1 + C_j e^{u\mu_j},$$

where C_j is derived by the boundary condition $\widehat{z}_j(t) = z_j$, yielding $C_j = (z_j - 1)e^{-t\mu_j}$, and thus $\widehat{z}_j(u) = 1 + (z_j - 1)e^{-\mu_j(t-u)}$. Upon substituting the solution for $\widehat{z}_j(u)$ in the equation of $\widehat{s}_j(u)$ in (59), we obtain the ODE

$$-\frac{\mathrm{d}\widehat{s}_j(u)}{\mathrm{d}u} + \alpha_j \widehat{s}_j(u) + (1 + (z_j - 1)e^{-\mu_j(t-u)})\beta_j(\widehat{\boldsymbol{s}}(u)) - 1 = 0, \tag{60}$$

with terminal condition $\widehat{s}_j(t) = s_j$. For later purposes, we rephrase the ODE in Eqn. (60) into an ODE subject to an initial condition. To that end, let $v = t_0 + t - u$ such that $t_0 \leqslant v \leqslant t$ and the ODE for $\widehat{s}_j(\cdot)$ becomes

$$\frac{\mathrm{d}\widehat{s}_j(t_0 + t - v)}{\mathrm{d}v} + \alpha_j \widehat{s}_j(t_0 + t - v) + (1 + (z_j - 1)e^{-\mu_j(v-t_0)}\beta_j(\widehat{\boldsymbol{s}}(t_0 + t - v)) - 1 = 0.$$

Upon defining $\tilde{s}_j(v) = \widehat{s}_j(t_0 + t - v)$, we have that $\tilde{s}_j(\cdot)$ satisfies Eqn. (9), with initial condition $\tilde{s}_j(t_0) = \widehat{s}_j(t) = s_j$.

We can now solve the characteristic equation of $\zeta_{t_0}(\cdot)$. Since the original PDE in Eqn. (58) is non-homogeneous, we know the solution $\zeta_{t_0}(t, \boldsymbol{s}, \boldsymbol{z})$ is not constant along characteristics, but evolves according to the right-hand side of (58). Therefore, if we set $\widehat{\zeta}_{t_0}(u) := \zeta_{t_0}(u, \widehat{\boldsymbol{s}}(u), \widehat{\boldsymbol{z}}(u))$ to be the solution restricted to the characteristics, then $\widehat{\zeta}_{t_0}(\cdot)$ satisfies

$$\frac{\partial \widehat{\zeta}_{t_0}(u)}{\partial u} = -\widehat{\zeta}_{t_0}(u) \sum_{j=1}^{d} \alpha_j \overline{\lambda}_j \widehat{s}_j(u),$$

subject to the initial condition

$$\widehat{\zeta}_{t_0}(t_0) = \zeta_{t_0}(t_0, \widehat{\boldsymbol{z}}(t_0), \widehat{\boldsymbol{s}}(t_0)) = \prod_{j=1}^{d} \widehat{z}_j(t_0)^{Q_{j,0}} \exp\big(-\widehat{s}_j(t_0)\lambda_{j,0}\big).$$

Solving this yields

$$\widehat{\zeta}_{t_0}(u) = \prod_{j=1}^{d} \widehat{z}_j(t_0)^{Q_{j,0}} \exp\Big(-\widehat{s}_j(t_0)\lambda_{j,0} - \alpha_j \overline{\lambda}_j \int_{t_0}^{u} \widehat{s}_j(v)\mathrm{d}v\Big).$$

Finally, the solution of the PDE is given at the endpoint of the characteristic $(t, \boldsymbol{s}, \boldsymbol{z})$, implying that $\zeta_{t_0}(t, \boldsymbol{s}, \boldsymbol{z}) = \widehat{\zeta}_{t_0}(t)$. Using the relation $\widehat{s}_j(t_0) = \tilde{s}_j(t)$, we obtain

$$\begin{aligned}\zeta_{t_0}(t, \boldsymbol{s}, \boldsymbol{z}) &= \prod_{j=1}^{d} \widehat{z}_j(t_0)^{Q_{j,0}} \exp\Big(- \widehat{s}_j(t_0)\lambda_{j,0} - \alpha_j\overline{\lambda}_j \int_0^t \widehat{s}_j(t_0 + t - u)\mathrm{d}u\Big)\\ &= \prod_{j=1}^{d} \widehat{z}_j(t_0)^{Q_{j,0}} \exp\Big(- \tilde{s}_j(t)\lambda_{j,0} - \alpha_j\overline{\lambda}_j \int_{t_0}^t \tilde{s}_j(u)\mathrm{d}u\Big),\end{aligned}$$

which finishes the proof. □

Proof of Theorem 2. The proof follows by conditioning on $\boldsymbol{Q}(t)$ and $\boldsymbol{\lambda}(t)$, and then applying Theorem 1 and techniques from its proof. By the tower property we have

$$\begin{aligned}&\mathbb{E}\Big[\prod_{i=1}^{d} y_i^{Q_i(t)} e^{-r_i\lambda_i(t)} z_i^{Q_i(t+\tau)} e^{-s_i\lambda_i(t+\tau)}\Big]\\ &= \mathbb{E}\Big[\prod_{i=1}^{d} y_i^{Q_i(t)} e^{-r_i\lambda_i(t)} \mathbb{E}\Big[\prod_{i=1}^{d} z_i^{Q_i(t+\tau)} e^{-s_i\lambda_i(t+\tau)} \,|\, \boldsymbol{Q}(t), \boldsymbol{\lambda}(t)\Big]\Big].\end{aligned} \tag{61}$$

The inner expectation can be derived from Theorem 1 and is given by

$$\mathbb{E}\Big[\prod_{i=1}^{d} z_i^{Q_i(t+\tau)} e^{-s_i\lambda_i(t+\tau)} \,|\, \boldsymbol{Q}(t), \boldsymbol{\lambda}(t)\Big] = \prod_{j=1}^{d} \widehat{z}_j(t)^{Q_j(t)} e^{-\tilde{s}_j(t+\tau)\lambda_j(t)} \exp\Big(- \overline{\lambda}_j\alpha_j \int_t^{t+\tau} \tilde{s}_j(u)\mathrm{d}u\Big),$$

where $\widehat{z}_j(\cdot)$ and $\tilde{s}_j(\cdot)$ satisfy Eqn. (15). Substituting this back into Eqn. (61) yields

$$\begin{aligned}&\mathbb{E}\Big[\prod_{i=1}^{d} y_i^{Q_i(t)} e^{-r_i\lambda_i(t)} \prod_{j=1}^{d} \widehat{z}_j(t)^{Q_j(t)} e^{-\tilde{s}_j(t+\tau)\lambda_j(t)} \exp\Big(- \overline{\lambda}_j\alpha_j \int_t^{t+\tau} \tilde{s}_j(u)\mathrm{d}u\Big)\Big]\\ &= \mathbb{E}\Big[\prod_{j=1}^{d} (y_j\widehat{z}_j(t))^{Q_j(t)} e^{-(r_j+\tilde{s}_j(t+\tau))\lambda_j(t)}\Big] \prod_{j=1}^{d} \exp\Big(- \overline{\lambda}_j\alpha_j \int_t^{t+\tau} \tilde{s}_j(u)\mathrm{d}u\Big)\\ &= \zeta(t, \boldsymbol{y} \odot \widehat{\boldsymbol{z}}(t), \boldsymbol{r} + \tilde{\boldsymbol{s}}(t+\tau)) \prod_{j=1}^{d} \exp\Big(- \overline{\lambda}_j\alpha_j \int_t^{t+\tau} \tilde{s}_j(u)\mathrm{d}u\Big).\end{aligned}$$

Applying Corollary 1 to the $\zeta(\cdot)$ term on the right-hand side, specifically Eqn. (10),

$$\zeta(t, \boldsymbol{y} \odot \widehat{\boldsymbol{z}}(t), \boldsymbol{r} + \tilde{\boldsymbol{s}}(t+\tau)) = \prod_{j=1}^{d} \exp\Big(- \overline{\lambda}_j\tilde{r}_j(t) - \overline{\lambda}_j\alpha_j \int_0^t \tilde{r}_j(v)\mathrm{d}v\Big),$$

where $\tilde{r}_j(\cdot)$ satisfies, for each $j \in [d]$, the ODE

$$\frac{\mathrm{d}\tilde{r}_j(v)}{\mathrm{d}v} + \alpha_j\tilde{r}_j(v) + \big(1 + (y_j\widehat{z}_j(t) - 1)e^{-\mu_j v}\big)\beta(\tilde{\boldsymbol{r}}(v)) - 1 = 0.$$

Since $\widehat{z}_j(t) = 1 + (z_j - 1)e^{-\mu_j\tau}$, substituting this into the ODE for $\tilde{r}_j(\cdot)$ and rearranging terms finishes the proof. □

Appendix B. Computations Related to Theorems 1 and 2

B.1. **Transform computations.** In this section, we provide the details behind taking the Laplace and $\boldsymbol{z}$-transform of Eqns. (55) and (56) respectively. First we show the Laplace transform, denoted by $\mathcal{L}(\cdot)$, of (55), which we restate here for convenience

$$\begin{aligned}
&\frac{\partial f(t,\boldsymbol{\nu},\boldsymbol{k})}{\partial t}-\sum_{j=1}^{d}\alpha_j\frac{\partial}{\partial \nu_j}\nu_j f(t,\boldsymbol{\nu},\boldsymbol{k})+\sum_{j=1}^{d}\alpha_j\overline{\lambda}_j\frac{\partial f(t,\boldsymbol{\nu},\boldsymbol{k})}{\partial \nu_j}\\
&=\sum_{j=1}^{d}\int_0^{\nu_d}\cdots\int_0^{\nu_1} y_j f(t,\boldsymbol{y},\boldsymbol{k}-\boldsymbol{e}_j)\frac{\partial^d}{\partial\nu_1\cdots\partial\nu_d}\mathbb{P}(\boldsymbol{B}_j\leqslant\boldsymbol{\nu}-\boldsymbol{y})\mathrm{d}y_1\cdots\mathrm{d}y_d\\
&+\sum_{j=1}^{d}\Big(f(t,\boldsymbol{\nu},\boldsymbol{k}+\boldsymbol{e}_j)(k_j+1)\mu_j-f(t,\boldsymbol{\nu},\boldsymbol{k})(k_j\mu_j+\nu_j)\Big),
\end{aligned}$$

and we introduce the shorthand notation

$$\begin{aligned}
\xi(t,\boldsymbol{s},\boldsymbol{k})&:=\mathcal{L}(f(t,\boldsymbol{\nu},\boldsymbol{k}))(\boldsymbol{s})\\
&=\int_0^\infty\cdots\int_0^\infty e^{-\boldsymbol{s}^\top\boldsymbol{\nu}}f(t,\boldsymbol{\nu},\boldsymbol{k})\,\mathrm{d}\nu_1\cdots\mathrm{d}\nu_d\equiv\int_{\boldsymbol{0}}^{\infty}e^{-\boldsymbol{s}^\top\boldsymbol{\nu}}f(t,\boldsymbol{\nu},\boldsymbol{k})\,\mathrm{d}\boldsymbol{\nu}.
\end{aligned}$$

We consider the term-by-term derivation in the equation that yields the transformed version as given in Eqn. (56). For the first term, it is clear that

$$\mathcal{L}\Big(\frac{\partial f(t,\boldsymbol{\nu},\boldsymbol{k})}{\partial t}\Big)(\boldsymbol{s})=\frac{\partial\xi(t,\boldsymbol{k},\boldsymbol{s})}{\partial t}.$$

For the second term, we need to show

$$-\mathcal{L}\Big(\sum_{j=1}^{d}\alpha_j\frac{\partial}{\partial\nu_j}\nu_j f(t,\boldsymbol{\nu},\boldsymbol{k})\Big)(\boldsymbol{s})=\sum_{j=1}^{d}\alpha_j s_j\frac{\partial}{\partial s_j}\xi(t,\boldsymbol{k},\boldsymbol{s}).$$

The argument of the Laplace transform is

$$\sum_{j=1}^{d}\alpha_j\frac{\partial}{\partial\nu_j}\nu_j f(t,\boldsymbol{\nu},\boldsymbol{k})=\sum_{j=1}^{d}\alpha_j f(t,\boldsymbol{\nu},\boldsymbol{k})+\sum_{j=1}^{d}\alpha_j\nu_j\frac{\partial f(t,\boldsymbol{\nu},\boldsymbol{k})}{\partial\nu_j}.$$

We then use the linearity of the Laplace transform and apply integration by parts to obtain

$$\begin{aligned}
&\sum_{j=1}^{d}\alpha_j\mathcal{L}\big(f(t,\boldsymbol{\nu},\boldsymbol{k})\big)(\boldsymbol{s})+\sum_{j=1}^{d}\alpha_j\mathcal{L}\Big(\nu_j\frac{\partial f(t,\boldsymbol{\nu},\boldsymbol{k})}{\partial\nu_j}\Big)(\boldsymbol{s})\\
&=\sum_{j=1}^{d}\alpha_j\int_{\boldsymbol{0}}^{\infty}e^{-\boldsymbol{s}^\top\boldsymbol{\nu}}f(t,\boldsymbol{\nu},\boldsymbol{k})\mathrm{d}\boldsymbol{\nu}+\sum_{j=1}^{d}\alpha_j\int_{\boldsymbol{0}}^{\infty}e^{-\boldsymbol{s}^\top\boldsymbol{\nu}}\nu_j\frac{\partial f(t,\boldsymbol{\nu},\boldsymbol{k})}{\partial\nu_j}\mathrm{d}\boldsymbol{\nu}\\
&=\sum_{j=1}^{d}\alpha_j\int_{\boldsymbol{0}}^{\infty}e^{-\boldsymbol{s}^\top\boldsymbol{\nu}}f(t,\boldsymbol{\nu},\boldsymbol{k})\mathrm{d}\boldsymbol{\nu}+\sum_{j=1}^{d}\alpha_j\big[\nu_j e^{-\boldsymbol{s}^\top\boldsymbol{\nu}}f(t,\boldsymbol{\nu},\boldsymbol{k})\big]_{\boldsymbol{0}}^{\infty}\\
&-\sum_{j=1}^{d}\alpha_j\int_{\boldsymbol{0}}^{\infty}\big(1-\nu_j s_j\big)e^{-\boldsymbol{s}^\top\boldsymbol{\nu}}f(t,\boldsymbol{\nu},\boldsymbol{k})\mathrm{d}\boldsymbol{\nu}
\end{aligned}$$

$$
\begin{aligned}
&= 0 + \sum_{j=1}^{d} \alpha_j s_j \int_{\mathbf{0}}^{\infty} \nu_j e^{-\boldsymbol{s}^\top \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d}\boldsymbol{\nu} \\
&= \sum_{j=1}^{d} \alpha_j s_j \int_{\mathbf{0}}^{\infty} -\frac{\partial}{\partial s_j} e^{-\boldsymbol{s}^\top \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d}\boldsymbol{\nu} = -\sum_{j=1}^{d} \alpha_j s_j \frac{\partial}{\partial s_j} \xi(t, \boldsymbol{k}, \boldsymbol{s}).
\end{aligned}
$$

For the third term, we need to show

$$
\sum_{j=1}^{d} \alpha_j \overline{\lambda}_j \mathcal{L}\Big(\frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_j}\Big)(\boldsymbol{s}) = \sum_{j=1}^{d} \alpha_j \overline{\lambda}_j s_j \xi(t, \boldsymbol{k}, \boldsymbol{s}).
$$

Using integration by parts and that $f(t, \boldsymbol{k}, \mathbf{0}) = 0$, we have

$$
\begin{aligned}
\mathcal{L}\Big(\frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_j}\Big)(\boldsymbol{s}) &= \int_{\mathbf{0}}^{\infty} e^{-\boldsymbol{s}^\top \boldsymbol{\nu}} \frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_j} \mathrm{d}\boldsymbol{\nu} \\
&= \big[e^{-\boldsymbol{s}^\top \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k})\big]_{\mathbf{0}}^{\infty} + s_j s_j \int_{\mathbf{0}}^{\infty} e^{-\boldsymbol{s}^\top \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d}\boldsymbol{\nu} = 0 + s_j \xi(t, \boldsymbol{k}, \boldsymbol{s}).
\end{aligned}
$$

For the fourth term, let us denote the probability density function of $\boldsymbol{B}_j = (B_{1j}, \ldots, B_{dj})^\top$ by $h_j(\cdot)$. We need to show

$$
\begin{aligned}
&\sum_{j=1}^{d} \mathcal{L}\Big(\int_{0}^{\nu_d} \cdots \int_{0}^{\nu_1} y_j f(t, \boldsymbol{y}, \boldsymbol{k} - \boldsymbol{e}_j) \frac{\partial^d}{\partial \nu_1 \cdots \partial \nu_d} \mathbb{P}(\boldsymbol{B}_j \leqslant \boldsymbol{\nu} - \boldsymbol{y}) \mathrm{d}y_1 \cdots \mathrm{d}y_d\Big) \\
&= \sum_{j=1}^{d} \mathcal{L}\Big(\int_{\mathbf{0}}^{\boldsymbol{\nu}} y_j h_j(\boldsymbol{\nu} - \boldsymbol{y}) f(t, \boldsymbol{y}, \boldsymbol{k} - \boldsymbol{e}_j) \mathrm{d}\boldsymbol{y}\Big) \\
&= -\sum_{j=1}^{d} \beta_j(\boldsymbol{s}) \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k} - \boldsymbol{e}_j)}{\partial s_j},
\end{aligned}
$$

with $\beta_j(\boldsymbol{s}) = \mathbb{E}[e^{-\boldsymbol{s}^\top \boldsymbol{B}_j}]$. To show that this holds, we need the property that relates convolutions with integration, which states that

$$
\int_{\mathbb{R}^d} (f * g)(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \Big(\int_{\mathbb{R}^d} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}\Big)\Big(\int_{\mathbb{R}^d} g(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}\Big),
$$

for given integrable functions f and g. Using this, we have

$$
\begin{aligned}
&-\sum_{j=1}^{d} \beta_j(\boldsymbol{s}) \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k} - \boldsymbol{e}_j)}{\partial s_j} \\
&\quad = -\sum_{j=1}^{d} \beta_j(\boldsymbol{s}) \int_{\mathbf{0}}^{\infty} \frac{\partial}{\partial s_j} e^{-\boldsymbol{s}^\top \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k} - \boldsymbol{e}_j) \mathrm{d}\boldsymbol{\nu} \\
&\quad = \sum_{j=1}^{d} \int_{\mathbf{0}}^{\infty} e^{-\boldsymbol{s}^\top \boldsymbol{\nu}} h_j(\boldsymbol{\nu}) \mathrm{d}\boldsymbol{\nu} \int_{\mathbf{0}}^{\infty} \nu_j e^{-\boldsymbol{s}^\top \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k} - \boldsymbol{e}_j) \mathrm{d}\boldsymbol{\nu} \\
&\quad = \sum_{j=1}^{d} \int_{\mathbf{0}}^{\infty} \int_{\mathbf{0}}^{\infty} e^{-\boldsymbol{s}^\top (\boldsymbol{\nu} - \boldsymbol{y})} h_j(\boldsymbol{\nu} - \boldsymbol{y}) y_j e^{-\boldsymbol{s}^\top \boldsymbol{y}} f(t, \boldsymbol{y}, \boldsymbol{k} - \boldsymbol{e}_j) \mathrm{d}\boldsymbol{y} \mathrm{d}\boldsymbol{\nu} \\
&\quad \overset{(\star)}{=} \sum_{j=1}^{d} \int_{\mathbf{0}}^{\infty} e^{-\boldsymbol{s}^\top \boldsymbol{\nu}} \int_{\mathbf{0}}^{\boldsymbol{\nu}} h_j(\boldsymbol{\nu} - \boldsymbol{y}) y_j f(t, \boldsymbol{y}, \boldsymbol{k} - \boldsymbol{e}_j) \mathrm{d}\boldsymbol{y} \mathrm{d}\boldsymbol{\nu}
\end{aligned}
$$

$$= \sum_{j=1}^{d} \mathcal{L}\Big(\int_0^{\nu_d} \cdots \int_0^{\nu_1} y_j f(t, \boldsymbol{y}, \boldsymbol{k} - \boldsymbol{e}_j) \frac{\partial^d}{\partial \nu_1 \cdots \partial \nu_d} \mathbb{P}(\boldsymbol{B}_j \leqslant \boldsymbol{\nu} - \boldsymbol{y}) \mathrm{d}y_1 \cdots \mathrm{d}y_d \Big),$$

where $(\star)$ holds because the non negativity $\mathbb{P}(\boldsymbol{B}_j \geqslant \boldsymbol{0}) = 1$ implies $h_j(\boldsymbol{\nu} - \boldsymbol{y}) = 0$ if $\boldsymbol{\nu} \geqslant \boldsymbol{y}$. The fifth term follows immediately by linearity since

$$\sum_{j=1}^{d} \mathcal{L}\big(f(t, \boldsymbol{\nu}, \boldsymbol{k} + \boldsymbol{e}_j)(k_j + 1)\mu_j\big)(\boldsymbol{s}) = \sum_{j=1}^{d} (k_j + 1)\mu_j \xi(t, \boldsymbol{s}, \boldsymbol{k} + \boldsymbol{e}_j).$$

Finally, the sixth term follows from the elementary computation

$$\begin{aligned} -\sum_{j=1}^{d} \mathcal{L}\big(f(t, \boldsymbol{\nu}, \boldsymbol{k})(k_j\mu_j + \nu_j)\big)(\boldsymbol{s}) &= -\sum_{j=1}^{d} k_j\mu_j \mathcal{L}\big(f(t, \boldsymbol{\nu}, \boldsymbol{k})(\boldsymbol{s}) - \sum_{j=1}^{d} k_j \nu_j \mathcal{L}\big(f(t, \boldsymbol{\nu}, \boldsymbol{k})\big)(\boldsymbol{s}) \\ &= -\sum_{j=1}^{d} k_j\mu_j \xi(t, \boldsymbol{s}, \boldsymbol{k}) + \sum_{j=1}^{d} \int_{\boldsymbol{0}}^{\infty} \frac{\partial}{\partial s_j} e^{-\boldsymbol{s}^\top \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d}\boldsymbol{\nu} \\ &= -\sum_{j=1}^{d} k_j\mu_j \xi(t, \boldsymbol{s}, \boldsymbol{k}) + \sum_{j=1}^{d} \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial s_j}. \end{aligned}$$

We can now derive Eqn. (57), where we use the shorthand notation

$$\zeta(t, \boldsymbol{s}, \boldsymbol{z}) = \mathcal{Z}\big(\xi(t, \boldsymbol{s}, \cdot)\big)(\boldsymbol{z}) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_d=0}^{\infty} z_1^{k_1} \cdots z_d^{k_d} \xi(t, \boldsymbol{s}, \boldsymbol{k}) \equiv \sum_{\boldsymbol{k} \in \mathbb{N}_0^d} \boldsymbol{z}^{\boldsymbol{k}} \xi(t, \boldsymbol{s}, \boldsymbol{k}),$$

with $\mathbb{N}_0^d = \{0, 1, 2, \dots\}^d$. As before, we take the term-by-term zeta transformation and show that we obtain Eqn. (58). The first and second terms are immediate by construction and linearity, since we have

$$\begin{aligned} \mathcal{Z}\Big(\frac{\partial \xi(t, \boldsymbol{s}, \cdot)}{\partial t}\Big)(\boldsymbol{z}) &= \frac{\partial \zeta(t, \boldsymbol{s}, \boldsymbol{z})}{\partial t}, \\ \sum_{j=1}^{d} \mathcal{Z}\Big(\alpha_j s_j \frac{\partial \xi(t, \boldsymbol{s}, \cdot)}{\partial s_j}\Big)(\boldsymbol{z}) &= \sum_{j=1}^{d} (\alpha_j s_j - 1) \frac{\partial \zeta(t, \boldsymbol{s}, \boldsymbol{z})}{\partial s_j}. \end{aligned}$$

We proceed by analyzing the third term, with the notation $\mathbb{N}_j^d = \{\boldsymbol{n} \in \mathbb{N}^d : n_j \geqslant 1\}$ for $j \in [d]$, with mild abuse of notation,

$$\begin{aligned} \sum_{j=1}^{d} \mathcal{Z}\Big(\frac{\partial \xi(t, \boldsymbol{s}, \cdot - \boldsymbol{e}_j)}{\partial s_j} \beta_j(\boldsymbol{s})\Big)(\boldsymbol{z}) &= \sum_{j=1}^{d} \sum_{\boldsymbol{k} \in \mathbb{N}_j^d} \boldsymbol{z}^{\boldsymbol{k}} \beta_j(\boldsymbol{s}) \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k} - \boldsymbol{e}_j)}{\partial s_j} \\ &= \sum_{j=1}^{d} \beta_j(\boldsymbol{s}) z_j \sum_{\boldsymbol{k} \in \mathbb{N}_0^d} \boldsymbol{z}^{\boldsymbol{k}} \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial s_j} = \sum_{j=1}^{d} \beta_j(\boldsymbol{s}) z_j \frac{\partial \zeta(t, \boldsymbol{s}, \boldsymbol{z})}{\partial s_j}. \end{aligned}$$

For the fourth and fifth term, using elementary computations,

$$\sum_{j=1}^{d} \mathcal{Z}\Big(\mu_j(k_j + 1)\xi(t, \boldsymbol{s}, \boldsymbol{k} + \boldsymbol{e}_j) - \mu_j k_j \xi(t, \boldsymbol{s}, \boldsymbol{k})\Big)(\boldsymbol{z})$$

$$
\begin{aligned}
&= \sum_{j=1}^{d} \mu_j \sum_{\boldsymbol{k}\in\mathbb{N}_0^d} (k_j+1)\boldsymbol{z}^{\boldsymbol{k}}\xi(t,\boldsymbol{s},\boldsymbol{k}+\boldsymbol{e}_j) - k_j\boldsymbol{z}^{\boldsymbol{k}}\xi(t,\boldsymbol{s},\boldsymbol{k}) \\
&= \sum_{j=1}^{d} \mu_j \sum_{\boldsymbol{k}\in\mathbb{N}_j^d} k_j\boldsymbol{z}^{\boldsymbol{k}-\boldsymbol{e}_j}\xi(t,\boldsymbol{s},\boldsymbol{k}) - k_j z_j\boldsymbol{z}^{\boldsymbol{k}-\boldsymbol{e}_j}\xi(t,\boldsymbol{s},\boldsymbol{k}) \\
&= \sum_{j=1}^{d} \mu_j(1-z_j)\frac{\partial}{\partial z_j}\sum_{\boldsymbol{k}\in\mathbb{N}_0^d} \boldsymbol{z}^{\boldsymbol{k}}\xi(t,\boldsymbol{s},\boldsymbol{k}) = -\sum_{j=1}^{d}\mu_j(z_j-1)\frac{\partial\zeta(t,\boldsymbol{s},\boldsymbol{z})}{\partial z_j}.
\end{aligned}
$$

Finally, the sixth term follows immediately from the definition since

$$
\sum_{j=1}^{d}\alpha_j\overline{\lambda}_j s_j\mathcal{Z}\Big(\xi(t,\boldsymbol{s},\cdot)\Big)(\boldsymbol{z}) = \zeta(t,\boldsymbol{s},\boldsymbol{z})\sum_{j=1}^{d}\alpha_j\overline{\lambda}_j s_j.
$$

B.2. **Joint moments: Computations.** In this section, we provide the details behind the derivation of the PDE to ODE as given in Eqns. (62) and (21). Since we are taking partial derivatives with respect to multiple variables, systematic bookkeeping is crucial. Some terms are straightforward to compute, so we focus on the ones that require careful attention. We first show the result and then provide details about the (relatively) complicated terms.

Differentiating Eqn. (19) $n_{\lambda_1},\ldots,n_{\lambda_d}$ times with respect to $s_1,\ldots,s_d$, respectively, and then substituting $\boldsymbol{s}=\boldsymbol{0}$, yields

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\Big[\prod_{i=1}^{d}\lambda_i(t)^{n_{\lambda_i}}z_i^{Q_i(t)}\Big] + \sum_{j=1}^{d} n_{\lambda_j}\alpha_j\mathbb{E}\Big[\prod_{i=1}^{d}\lambda_i(t)^{n_{\lambda_i}}z_i^{Q_i(t)}\Big] \\
&\quad - \sum_{l=1}^{d} n_{\lambda_l}\sum_{j=1}^{d}\mathbb{E}\big[B_{lj}\big]\mathbb{E}\Big[z_j\lambda_j(t)\prod_{i=1}^{d}\lambda_i(t)^{n_{\lambda_i}-\mathbf{1}_{\{i=l\}}}z_i^{Q_i(t)}\Big] \\
&\quad + \sum_{j=1}^{d}\mu_j(z_j-1)\mathbb{E}\Big[Q_j(t)\prod_{i=1}^{d}\lambda_i(t)^{n_{\lambda_i}}z_i^{Q_i(t)-\mathbf{1}_{\{i=j\}}}\Big] \qquad (62)\\
&= \sum_{j=1}^{d}(z_j-1)\mathbb{E}\Big[\lambda_j(t)\prod_{i=1}^{d}\lambda_i(t)^{n_{\lambda_i}}z_i^{Q_i(t)}\Big] + \sum_{j=1}^{d}\alpha_j\overline{\lambda}_j n_{\lambda_j}\mathbb{E}\Big[\prod_{i=1}^{d}\lambda_i(t)^{n_{\lambda_i}-\mathbf{1}_{\{i=j\}}}z_i^{Q_i(t)}\Big] \\
&\quad + \sum_{j=1}^{d}\sum_{m_1=0}^{n_{\lambda_1}}\cdots\sum_{m_d=0}^{n_{\lambda_d}}\mathbf{1}_{\{m\leqslant n_\lambda-2\}}\prod_{k=1}^{d}\binom{n_{\lambda_k}}{m_k}\mathbb{E}\Big[z_j\prod_{i=1}^{d}B_{ij}^{n_{\lambda_i}-m_i}\lambda_i(t)^{m_i+\mathbf{1}_{\{i=j\}}}z_i^{Q_i(t)}\Big],
\end{aligned}
$$

where $m=\sum_{i=1}^{d}m_i$ and we collected the $\mathbb{E}[B_{ij}]$ combinations of first order on the left-hand side and higher orders on the right-hand side. Then take Eqn. (62) and differentiate $n_{Q_1},\ldots,n_{Q_d}$ times with respect to $z_1,\ldots,z_d$, respectively, and substitute $\boldsymbol{z}=\boldsymbol{1}$. After elementary calculus,

$$
\frac{\mathrm{d}}{\mathrm{d}t}\psi_t(\boldsymbol{n_\lambda},\boldsymbol{n_Q}) + \sum_{j=1}^{d}\big(n_{\lambda_j}(\alpha_j-\mathbb{E}\big[B_{jj}\big]) + n_{Q_j}\mu_j\big)\psi_t(\boldsymbol{n_\lambda},\boldsymbol{n_Q})
$$

$$
\begin{aligned}
&= \sum_{j=1}^{d} \sum_{\substack{i=1 \\ i \neq j}}^{d} n_{\lambda_i} \mathbb{E}\big[B_{ij}\big] \psi_t(\boldsymbol{n_\lambda} - \boldsymbol{e}_i + \boldsymbol{e}_j, \boldsymbol{n_Q}) + \sum_{j=1}^{d} n_{Q_j} \psi_t(\boldsymbol{n_\lambda} + \boldsymbol{e_j}, \boldsymbol{n_Q} - \boldsymbol{e}_j) \qquad (63)\\
&\quad + \sum_{j=1}^{d} \alpha_j \overline{\lambda}_j n_{\lambda_j} \psi_t(\boldsymbol{n_\lambda} - \boldsymbol{e}_j, \boldsymbol{n_Q}) + \sum_{i=1}^{d} \sum_{j=1}^{d} n_{\lambda_i} n_{Q_j} \mathbb{E}\big[B_{ij}\big] \psi_t(\boldsymbol{n_\lambda} - \boldsymbol{e}_i + \boldsymbol{e}_j, \boldsymbol{n_Q} - \boldsymbol{e}_j)\\
&\quad + \sum_{j=1}^{d} \sum_{m_1=0}^{n_{\lambda_1}} \cdots \sum_{m_d=0}^{n_{\lambda_d}} \mathbf{1}_{\{m \leqslant n_\lambda - 2\}} \prod_{k=1}^{d} \binom{n_{\lambda_k}}{m_k} \Big\{ n_{Q_j} \prod_{i=1}^{d} \mathbb{E}\big[B_{ij}^{n_{\lambda_i} - m_i}\big] \psi_t(\boldsymbol{m} + \boldsymbol{e}_j, \boldsymbol{n_Q} - \boldsymbol{e}_j)\\
&\qquad + \prod_{i=1}^{d} \mathbb{E}\big[B_{ij}^{n_{\lambda_i} - m_i}\big] \psi_t(\boldsymbol{m} + \boldsymbol{e}_j, \boldsymbol{n_Q}) \Big\}.
\end{aligned}
$$

To obtain the ODE in Eqn. (62), the starting point is Eqn. (19). Differentiate $n_{\lambda_1}, \dots, n_{\lambda_d}$ times with respect to $s_1, \dots, s_d$ respectively, and then substitute $\boldsymbol{s} = \boldsymbol{0}$. The terms that are not immediate to compute are those where we need to apply the product rule repeatedly. Consider the computation of

$$
\frac{\partial^{n_{\lambda_1}} \cdots \partial^{n_{\lambda_d}}}{\partial s_1^{n_{\lambda_1}} \cdots \partial s_d^{n_{\lambda_d}}} \sum_{j=1}^{d} \alpha_j s_j \mathbb{E}\big[\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_n^{Q_n(t)}\big].
$$

We first focus on differentiation with respect to the first component, yielding

$$
\begin{aligned}
&\frac{\partial^{n_{\lambda_1}}}{\partial s_1^{n_{\lambda_1}}} \sum_{j=1}^{d} \alpha_j s_j \mathbb{E}\big[\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_n^{Q_n(t)}\big]\\
&= \frac{\partial^{n_{\lambda_1}-1}}{\partial s_1^{n_{\lambda_1}-1}} \alpha_1 \mathbb{E}\big[\lambda_1(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_n^{Q_n(t)}\big] - \frac{\partial^{n_{\lambda_1}-1}}{\partial s_1^{n_{\lambda_1}-1}} \sum_{j=1}^{d} \alpha_j s_j \mathbb{E}\big[\lambda_1(t) \lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_n^{Q_n(t)}\big]\\
&= -2 \frac{\partial^{n_{\lambda_1}-2}}{\partial s_1^{n_{\lambda_1}-2}} \alpha_1 \mathbb{E}\big[\lambda_1(t)^2 e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_n^{Q_n(t)}\big] + \frac{\partial^{n_{\lambda_1}-2}}{\partial s_1^{n_{\lambda_1}-2}} \sum_{j=1}^{d} \alpha_j s_j \mathbb{E}\big[\lambda_1(t)^2 \lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_n^{Q_n(t)}\big]\\
&\ \vdots\\
&= n_{\lambda_1} (-1)^{n_{\lambda_1}-1} \alpha_1 \mathbb{E}\big[\lambda_1(t)^{n_{\lambda_1}} e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_n^{Q_n(t)}\big] + \sum_{j=1}^{d} \alpha_j s_j \mathbb{E}\big[\lambda_1(t)^{n_{\lambda_1}} \lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_n^{Q_n(t)}\big].
\end{aligned}
$$

Note that all the terms in the latter sum vanish when we substitute $\boldsymbol{s} = \boldsymbol{0}$. An analogous expression holds for the other components. If we now combine the differentiation with respect to all components and substitute $\boldsymbol{s} = \boldsymbol{0}$, we have

$$
\frac{\partial^{n_{\lambda_1}} \cdots \partial^{n_{\lambda_d}}}{\partial s_1^{n_{\lambda_1}} \cdots \partial s_d^{n_{\lambda_d}}} \sum_{j=1}^{d} \alpha_j s_j \mathbb{E}\big[\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_n^{Q_n(t)}\big] = \sum_{j=1}^{d} n_{\lambda_j} \alpha_j \mathbb{E}\Big[\prod_{i=1}^{d} \lambda_i(t)^{n_{\lambda_i}} z_i^{Q_i(t)}\Big].
$$

Another, more complicated, term we need to compute is

$$
\frac{\partial^{n_{\lambda_1}} \cdots \partial^{n_{\lambda_d}}}{\partial s_1^{n_{\lambda_1}} \cdots \partial s_d^{n_{\lambda_d}}} \sum_{j=1}^{d} z_j \beta_j(\boldsymbol{s}) \mathbb{E}\big[\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_n^{Q_n(t)}\big],
$$

with $\beta_j(\boldsymbol{s}) = \mathbb{E}\big[e^{-\boldsymbol{s}^\top \boldsymbol{B}_j}\big]$. It is clear that taking higher order derivatives means that we have to successively apply the product rule. Moreover, since we are taking partial derivatives with respect to multiple components, we obtain a large number of cross terms. Let us focus on the first component, which yields

$$
\begin{aligned}
&\frac{\partial^{n_{\lambda_1}}}{\partial s_1^{n_{\lambda_1}}} \sum_{j=1}^d z_j \beta_j(\boldsymbol{s}) \mathbb{E}\big[\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^d z_n^{Q_n(t)}\big] \\
&= (-1)^1 \sum_{j=1}^d z_j \frac{\partial^{n_{\lambda_1}-1}}{\partial s_1^{n_{\lambda_1}-1}} \mathbb{E}\big[B_{1j} e^{-\boldsymbol{s}^\top \boldsymbol{B}_j}\big] \mathbb{E}\big[\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^d z_n^{Q_n(t)}\big] \\
&\quad + (-1)^1 \sum_{j=1}^d z_j \frac{\partial^{n_{\lambda_1}-1}}{\partial s_1^{n_{\lambda_1}-1}} \mathbb{E}\big[e^{-\boldsymbol{s}^\top \boldsymbol{B}_j}\big] \mathbb{E}\big[\lambda_1(t)\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^d z_n^{Q_n(t)}\big] \\
&= (-1)^2 \sum_{j=1}^d z_j \frac{\partial^{n_{\lambda_1}-2}}{\partial s_1^{n_{\lambda_1}-2}} \mathbb{E}\big[B_{1j}^2 e^{-\boldsymbol{s}^\top \boldsymbol{B}_j}\big] \mathbb{E}\big[\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^d z_n^{Q_n(t)}\big] \\
&\quad + 2(-1)^2 \sum_{j=1}^d z_j \frac{\partial^{n_{\lambda_1}-2}}{\partial s_1^{n_{\lambda_1}-2}} \mathbb{E}\big[B_{1j} e^{-\boldsymbol{s}^\top \boldsymbol{B}_j}\big] \mathbb{E}\big[\lambda_1(t)\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^d z_n^{Q_n(t)}\big] \\
&\quad + (-1)^2 \sum_{j=1}^d z_j \frac{\partial^{n_{\lambda_1}-2}}{\partial s_1^{n_{\lambda_1}-2}} \mathbb{E}\big[e^{-\boldsymbol{s}^\top \boldsymbol{B}_j}\big] \mathbb{E}\big[\lambda_1(t)^2\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^d z_n^{Q_n(t)}\big] \\
&\vdots \\
&= n_{\lambda_1} (-1)^{n_{\lambda_1}} \sum_{j=1}^d z_j \mathbb{E}\big[B_{1j} e^{-\boldsymbol{s}^\top \boldsymbol{B}_j}\big] \mathbb{E}\big[\lambda_1(t)^{n_{\lambda_1}-1}\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^d z_n^{Q_n(t)}\big] \\
&\quad + (-1)^{n_{\lambda_1}} \sum_{j=1}^d z_j \mathbf{1}_{\{n_{\lambda_1} \geqslant 2\}} \sum_{m_1=0}^{n_{\lambda_1}-2} \binom{n_{\lambda_1}}{m_1} \mathbb{E}\big[B_{1j}^{n_{\lambda_1}-m_1}\big] \mathbb{E}\big[\lambda_1(t)^{m}\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^d z_n^{Q_n(t)}\big] \\
&\quad + (-1)^{n_{\lambda_1}} \sum_{j=1}^d z_j \mathbb{E}\big[\lambda_1(t)^{n_{\lambda_1}}\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^d z_n^{Q_n(t)}\big],
\end{aligned}
$$

since the number of terms is doubled in every step of the derivation. The computation for the other components is entirely analogous. Upon taking the joint derivative and substituting $\boldsymbol{s} = \boldsymbol{0}$, we obtain

$$
\begin{aligned}
&\frac{\partial^{n_{\lambda_1}} \cdots \partial^{n_{\lambda_d}}}{\partial s_1^{n_{\lambda_1}} \cdots \partial s_d^{n_{\lambda_d}}} \sum_{j=1}^d z_j \beta_j(\boldsymbol{s}) \mathbb{E}\big[\lambda_j(t) e^{-\boldsymbol{s}^\top \boldsymbol{\lambda}(t)} \prod_{n=1}^d z_n^{Q_n(t)}\big] \\
&= \sum_{l=1}^d n_{\lambda_l} \sum_{j=1}^d \mathbb{E}\big[B_{lj}\big] \mathbb{E}\Big[z_j \lambda_j(t) \prod_{i=1}^d \lambda_i(t)^{n_{\lambda_i} - \mathbf{1}_{\{i=l\}}} z_i^{Q_i(t)}\Big] + \sum_{j=1}^d z_j \mathbb{E}\Big[\lambda_j(t) \prod_{i=1}^d \lambda_i(t)^{n_{\lambda_i}} z_i^{Q_i(t)}\Big] \\
&\quad + \sum_{j=1}^d \sum_{m_1=0}^{n_{\lambda_1}} \cdots \sum_{m_d=0}^{n_{\lambda_d}} \mathbf{1}_{\{m \leqslant n_\lambda - 2\}} \prod_{k=1}^d \binom{n_{\lambda_k}}{m_k} \mathbb{E}\Big[z_j \prod_{i=1}^d B_{ij}^{n_{\lambda_i}-m_i} \lambda_i(t)^{m_i + \mathbf{1}_{\{i=j\}}} z_i^{Q_i(t)}\Big],
\end{aligned}
$$

where $m = m_1 + \cdots m_d$.

We now focus on the terms to obtain the ODE in Eqn. (21). The starting point is Eqn. (62), which we differentiate $n_{Q_1}, \ldots, n_{Q_d}$ times with respect to $z_1, \ldots, z_d$ respectively and substitute $\boldsymbol{z} = \mathbf{1}$. There are multiple terms in (62) that require the product rule when differentiating. We consider one such term and take the appropriate derivative, i.e.,

$$\frac{\partial^{n_{Q_1}} \cdots \partial^{n_{Q_d}}}{\partial z_1^{n_{Q_1}} \cdots \partial z_d^{n_{Q_d}}} \sum_{j=1}^{d} \mu_j (z_j - 1) \mathbb{E}\Big[Q_j(t) \prod_{i=1}^{d} \lambda_i(t)^{n_{\lambda_i}} z_i^{Q_i(t) - \mathbf{1}_{\{i=j\}}} \Big].$$

Again, we focus on differentiation with respect to the first component, which yields

$$\begin{aligned}
&\frac{\partial^{n_{Q_1}}}{\partial z_1^{n_{Q_1}}} \sum_{j=1}^{d} \mu_j (z_j - 1) \mathbb{E}\Big[Q_j(t) \prod_{i=1}^{d} \lambda_i(t)^{n_{\lambda_i}} z_i^{Q_i(t) - \mathbf{1}_{\{i=j\}}} \Big] \\
&= \frac{\partial^{n_{Q_1}-1}}{\partial z_1^{n_{Q_1}-1}} \mu_1 \mathbb{E}\Big[Q_1(t) \prod_{i=1}^{d} \lambda_i(t)^{n_{\lambda_i}} z_i^{Q_i(t) - \mathbf{1}_{\{i=1\}}} \Big] \\
&\quad + \frac{\partial^{n_{Q_1}-1}}{\partial z_1^{n_{Q_1}-1}} \mu_1 (z_1 - 1) \mathbb{E}\Big[Q_1(t)(Q_1(t) - 1) \prod_{i=1}^{d} \lambda_i(t)^{n_{\lambda_i}} z_i^{Q_i(t) - 2\mathbf{1}_{\{i=1\}}} \Big] \\
&\quad + \sum_{j=2}^{d} \frac{\partial^{n_{Q_1}-1}}{\partial z_1^{n_{Q_1}-1}} \mu_j (z_j - 1) \mathbb{E}\Big[Q_j(t) Q_1(t) \prod_{i=1}^{d} \lambda_i(t)^{n_{\lambda_i}} z_i^{Q_i(t) - \mathbf{1}_{\{i=j\}} - \mathbf{1}_{\{i=1\}}} \Big] \\
&= 2 \frac{\partial^{n_{Q_1}-2}}{\partial z_1^{n_{Q_1}-2}} \mu_1 \mathbb{E}\Big[Q_1(t)(Q_1(t) - 1) \prod_{i=1}^{d} \lambda_i(t)^{n_{\lambda_i}} z_i^{Q_i(t) - 2\mathbf{1}_{\{i=1\}}} \Big] \\
&\quad + \frac{\partial^{n_{Q_1}-2}}{\partial z_1^{n_{Q_1}-2}} \mu_1 (z_1 - 1) \mathbb{E}\Big[Q_1(t)(Q_1(t) - 1)(Q_1(t) - 2) \prod_{i=1}^{d} \lambda_i(t)^{n_{\lambda_i}} z_i^{Q_i(t) - 3\mathbf{1}_{\{i=1\}}} \Big] \\
&\quad \sum_{j=2}^{d} \frac{\partial^{n_{Q_1}-2}}{\partial z_1^{n_{Q_1}-2}} \mu_j (z_j - 1) \mathbb{E}\Big[Q_j(t) Q_1(t)(Q_1(t) - 1) \prod_{i=1}^{d} \lambda_i(t)^{n_{\lambda_i}} z_i^{Q_i(t) - \mathbf{1}_{\{i=j\}} - 2\mathbf{1}_{\{i=1\}}} \Big] \\
&\vdots \\
&= n_{Q_1} \mu_1 \mathbb{E}\Big[Q_1(t)^{[n_{Q_1}]} \prod_{i=1}^{d} \lambda_i(t)^{n_{\lambda_i}} z_i^{Q_i(t) - n_{Q_1} \mathbf{1}_{\{i=1\}}} \Big],
\end{aligned}$$

where we substituted $\boldsymbol{z} = \mathbf{1}$ in the last step, canceling out all the terms that contain the factor $(z_j - 1)$. We can compute the derivatives with respect to other components in a similar manner, which results in

$$\begin{aligned}
&\frac{\partial^{n_{Q_1}} \cdots \partial^{n_{Q_d}}}{\partial z_1^{n_{Q_1}} \cdots \partial z_d^{n_{Q_d}}} \sum_{j=1}^{d} \mu_j (z_j - 1) \mathbb{E}\Big[Q_j(t) \prod_{i=1}^{d} \lambda_i(t)^{n_{\lambda_i}} z_i^{Q_i(t) - \mathbf{1}_{\{i=j\}}} \Big] \\
&= \sum_{j=1}^{d} n_{Q_j} \mathbb{E}\Big[\prod_{i=1}^{d} \lambda_i^{n_{\lambda_i}} Q_i(t)^{[n_{Q_i}]} \Big] = \sum_{j=1}^{d} n_{Q_j} \psi_t(\boldsymbol{n_Q}, \boldsymbol{n_\lambda}).
\end{aligned}$$

Appendix C. First and Second Order Transient and Stationary Moments

In this appendix we present an illustration concerning moments of order $n \in \{1, 2\}$. We introduce the relevant objects along the way, starting with the matrix $\mathbb{E}[\boldsymbol{B}] = (\mathbb{E}[B_{ij}])_{i,j \in [d]}$

and the diagonal matrices

$$\boldsymbol{D}_\alpha := \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_d), \quad \boldsymbol{D}_\mu := \mathrm{diag}(\mu_1, \mu_2, \ldots, \mu_d).$$

C.1. **Transient moments.** We focus on the transient moments $\psi_t(\boldsymbol{n_Q}, \boldsymbol{n_\lambda})$, where now $\boldsymbol{n_Q} = (n_{Q_1}, \ldots, n_{Q_d})$ and $\boldsymbol{n_\lambda} = (n_{\lambda_1}, \ldots, n_{\lambda_d})$. We separately consider the cases $n = 1$ and $n = 2$. To describe the joint moments of equal order in vector/matrix-form, it turns out that for $n = 1$ we need a stacked *vector*, and for $n = 2$ a stacked *matrix*, introduced in detail below.

For $n = 1$ we define the stacked vector

$$\boldsymbol{\Sigma}_t^{(1)} := \Big(\mathbb{E}[\boldsymbol{\lambda}(t)], \mathbb{E}[\boldsymbol{Q}(t)]\Big)^\top. \tag{64}$$

For each entry of the vector, we use Eqn. (21) to obtain the vector-valued ODEs

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\boldsymbol{\lambda}(t)] &= \big(\mathbb{E}[\boldsymbol{B}] - \boldsymbol{D}_\alpha\big)\mathbb{E}[\boldsymbol{\lambda}(t)] + \boldsymbol{L}^{(0,1)}, \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\boldsymbol{Q}(t)] &= -\boldsymbol{D}_\mu\mathbb{E}[\boldsymbol{Q}(t)] + \mathbb{E}[\boldsymbol{\lambda}(t)], \end{aligned} \tag{65}$$

where $\boldsymbol{L}^{(0,1)} = \big(\alpha_1\overline{\lambda}_1, \alpha_2\overline{\lambda}_2, \ldots, \alpha_d\overline{\lambda}_d\big)^\top$. It is directly verified that (65) is solved by

$$\begin{aligned} \mathbb{E}[\boldsymbol{\lambda}(t)] &= e^{t(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_\alpha)}\overline{\boldsymbol{\lambda}} + \int_0^t e^{(t-s)(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_\alpha)}\mathrm{d}s\,(\boldsymbol{\alpha}\odot\overline{\boldsymbol{\lambda}}) \\ &= e^{t(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_\alpha)}\overline{\boldsymbol{\lambda}} + (\mathbb{E}[\boldsymbol{B}] - \boldsymbol{D}_\alpha)^{-1}\big(e^{t(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_\alpha)} - \boldsymbol{I}\big)(\boldsymbol{\alpha}\odot\overline{\boldsymbol{\lambda}}), \\ \mathbb{E}[\boldsymbol{Q}(t)] &= \int_0^t e^{-(t-s)\boldsymbol{D}_\mu}\mathbb{E}[\boldsymbol{\lambda}(s)]\mathrm{d}s. \end{aligned} \tag{66}$$

We proceed with $n = 2$. As indicated at the start of this subsection, in this case we should work with a stacked matrix. To this end, define

$$\begin{aligned} \mathbb{E}\big[\boldsymbol{Q}(t)^{[2]}\big] &:= \mathbb{E}\big[\boldsymbol{Q}(t)\boldsymbol{Q}(t)^\top\big] - \mathrm{diag}(\mathbb{E}[\boldsymbol{Q}(t)]) \\ &= \mathbb{E}\begin{bmatrix} Q_1(t)^{[2]} & Q_1(t)Q_2(t) & \cdots & Q_1(t)Q_d(t) \\ Q_2(t)Q_1(t) & Q_2(t)^{[2]} & \cdots & Q_2(t)Q_d(t) \\ \vdots & \vdots & \ddots & \vdots \\ Q_d(t)Q_1(t) & Q_d(t)Q_2(t) & \cdots & Q_d(t)^{[2]} \end{bmatrix}, \end{aligned}$$

and we also consider the objects $\mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{Q}(t)^\top\big]$ and $\mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{\lambda}(t)^\top\big]$, which are all $d\times d$-matrices. In addition, we define the stacked matrix $\boldsymbol{\Sigma}_t^{(2)}$ given by

$$\boldsymbol{\Sigma}_t^{(2)} := \mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{\lambda}(t)^\top\big] \oplus \mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{Q}(t)^\top\big] \oplus \mathbb{E}\big[\boldsymbol{Q}(t)^{[2]}\big], \tag{67}$$

where $\oplus$ indicates the direct sum, so that $\boldsymbol{\Sigma}_t^{(2)}$ is a $3d \times 3d$-matrix. For each entry of a submatrix, we derive its associated ODE from Eqn. (21) which we combine into matrix-valued ODEs. We thus find the matrix-valued ODEs

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{\lambda}(t)^\top\big] &= \big(\mathbb{E}[\boldsymbol{B}] - \boldsymbol{D}_\alpha\big)\mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{\lambda}(t)^\top\big] + \mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{\lambda}(t)^\top\big]\big(\mathbb{E}[\boldsymbol{B}] - \boldsymbol{D}_\alpha\big)^\top \\ &\quad + \mathbb{E}\big[\boldsymbol{B}\,\mathrm{diag}\big(\mathbb{E}[\boldsymbol{\lambda}(t)]\big)\boldsymbol{B}^\top\big] + \boldsymbol{D}_\alpha\big(\bar{\boldsymbol{\lambda}}\mathbb{E}[\boldsymbol{\lambda}(t)]^\top\big) + \big(\mathbb{E}[\boldsymbol{\lambda}(t)]\bar{\boldsymbol{\lambda}}^\top\big)\boldsymbol{D}_\alpha, \end{aligned}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{Q}(t)^\top\big] = \big(\mathbb{E}[\boldsymbol{B}] - \boldsymbol{D}_\alpha\big)\mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{Q}(t)^\top\big] - \mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{Q}(t)^\top\big]\boldsymbol{D}_\mu + \mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{\lambda}(t)^\top\big] \qquad (68)$$
$$+ (\boldsymbol{\alpha}\odot\bar{\boldsymbol{\lambda}})\mathbb{E}[\boldsymbol{Q}(t)]^\top + \mathbb{E}[\boldsymbol{B}]\operatorname{diag}\big(\mathbb{E}[\boldsymbol{\lambda}(t)]\big),$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[\boldsymbol{Q}(t)^{[2]}\big] = -\boldsymbol{D}_\mu\mathbb{E}\big[\boldsymbol{Q}(t)^{[2]}\big] - \mathbb{E}\big[\boldsymbol{Q}(t)^{[2]}\big]\boldsymbol{D}_\mu + \mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{Q}(t)^\top\big] + \big(\mathbb{E}\big[\boldsymbol{\lambda}(t)\boldsymbol{Q}(t)^\top\big]\big)^\top.$$

We end our account of the transient moments with a series of brief remarks. The ODEs for $\boldsymbol{\Sigma}_t^{(1)}$ and $\boldsymbol{\Sigma}_t^{(2)}$ are related to those derived in Lemmas 1 and 3 of [7]. Concretely, the solution for the first moment $\mathbb{E}[\boldsymbol{\lambda}(t)]$ agrees with Eqn. (8) in [7]. Furthermore, by taking the limit $\boldsymbol{\mu}\downarrow\boldsymbol{0}$ in our expression for $\boldsymbol{Q}(t)$, we obtain

$$\mathbb{E}\big[\boldsymbol{N}(t)\big] = \big(\mathbb{E}[\boldsymbol{B}] - \boldsymbol{D}_\alpha\big)^{-1}\big(e^{t\,(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_\alpha)} - I\big)\overline{\boldsymbol{\lambda}} +$$
$$\big(\mathbb{E}[\boldsymbol{B}] - \boldsymbol{D}_\alpha\big)^{-2}\big(e^{t\,(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_\alpha)} - I\big)(\boldsymbol{\alpha}\odot\overline{\boldsymbol{\lambda}}) + t\,\big(\mathbb{E}[\boldsymbol{B}] - \boldsymbol{D}_\alpha\big)^{-1}(\boldsymbol{\alpha}\odot\overline{\boldsymbol{\lambda}}), \qquad (69)$$

which agrees with the result in Eqn. (10) in [7]. Regarding the second order moments, upon taking $\boldsymbol{\mu}\downarrow 0$ in Eqn. (68) we recover the expressions in Lemma 3 of [7] (where it is noted that an elementary conversions needs to be performed, as we work with $\mathbb{E}\big[\boldsymbol{N}(t)^{[2]}\big]$ and [7] with $\mathbb{E}\big[\boldsymbol{N}(t)\boldsymbol{N}(t)^\top\big]$).

C.2. **Stationary moments.** We continue by considering the joint stationary moments of order at most 2. We adopt the notation used in Subsection C.1.

For order $n = 1$, define the stationary version of Eqn. (64):

$$\boldsymbol{\Sigma}^{(1)} := \lim_{t\to\infty}\boldsymbol{\Sigma}_t^{(1)} = \Big(\mathbb{E}[\boldsymbol{\lambda}], \mathbb{E}[\boldsymbol{Q}]\Big)^\top. \qquad (70)$$

We derive from Eqn. (23) that the elements of this stacked vector satisfy

$$\mathbb{E}[\boldsymbol{\lambda}] = -\big(\mathbb{E}[\boldsymbol{B}] - \boldsymbol{D}_\alpha\big)^{-1}\boldsymbol{L}^{(0,1)}, \qquad \mathbb{E}[\boldsymbol{Q}] = \boldsymbol{D}_\mu^{-1}\mathbb{E}[\boldsymbol{\lambda}]. \qquad (71)$$

For order $n = 2$, we define the stacked matrix

$$\boldsymbol{\Sigma}^{(2)} := \lim_{t\to\infty}\boldsymbol{\Sigma}_t^{(2)} = \mathbb{E}[\boldsymbol{\lambda}\boldsymbol{\lambda}^\top] \oplus \mathbb{E}[\boldsymbol{\lambda}\boldsymbol{Q}^\top] \oplus \mathbb{E}[\boldsymbol{Q}^{[2]}], \qquad (72)$$

From the procedure followed for the transient moments, in combination with Eqn. (23), we conclude that, with $\mathbb{E}[\boldsymbol{\lambda}]$ and $\mathbb{E}[\boldsymbol{Q}]$ given above,

$$0 = \big(\mathbb{E}[\boldsymbol{B}] - \boldsymbol{D}_\alpha\big)\mathbb{E}[\boldsymbol{\lambda}\boldsymbol{\lambda}^\top] + \mathbb{E}[\boldsymbol{\lambda}\boldsymbol{\lambda}^\top]\big(\mathbb{E}[\boldsymbol{B}] - \boldsymbol{D}_\alpha\big)^\top + \mathbb{E}\big[\boldsymbol{B}\operatorname{diag}\big(\mathbb{E}[\boldsymbol{\lambda}]\big)\boldsymbol{B}^\top\big]$$
$$+ \boldsymbol{D}_\alpha\big(\bar{\boldsymbol{\lambda}}\mathbb{E}[\boldsymbol{\lambda}]^\top\big) + \big(\mathbb{E}[\boldsymbol{\lambda}]\bar{\boldsymbol{\lambda}}^\top\big)\boldsymbol{D}_\alpha,$$
$$0 = \big(\mathbb{E}[\boldsymbol{B}] - \boldsymbol{D}_\alpha\big)\mathbb{E}[\boldsymbol{\lambda}\boldsymbol{Q}^\top] - \mathbb{E}[\boldsymbol{\lambda}\boldsymbol{Q}^\top]\boldsymbol{D}_\mu + \mathbb{E}[\boldsymbol{\lambda}\boldsymbol{\lambda}^\top] + (\boldsymbol{\alpha}\odot\bar{\boldsymbol{\lambda}})\mathbb{E}[\boldsymbol{Q}]^\top + \mathbb{E}[\boldsymbol{B}]\operatorname{diag}\big(\mathbb{E}[\boldsymbol{\lambda}]\big)^\top,$$
$$0 = -\boldsymbol{D}_\mu\mathbb{E}[\boldsymbol{Q}^{[2]}] - \mathbb{E}[\boldsymbol{Q}^{[2]}]\boldsymbol{D}_\mu + \mathbb{E}[\boldsymbol{\lambda}\boldsymbol{Q}^\top] + \big(\mathbb{E}[\boldsymbol{\lambda}\boldsymbol{Q}^\top]\big)^\top. \qquad (73)$$

The matrix-valued equations in (73) are all *Sylvester equations*, i.e., equations of the form

$$\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{B} = \boldsymbol{C}, \qquad (74)$$

for known matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, with the matrix $\boldsymbol{X}$ being unknown. It is a known result that a unique solution for $\boldsymbol{X}$ exists if and only $\boldsymbol{A}$ and $-\boldsymbol{B}$ do not share any eigenvalue.

We conclude this subsection with two results on higher order stationary moments. By applying Eqn. (23), we can obtain expressions for the moments

$$\psi(\mathbf{0}, \boldsymbol{n_\lambda}) = \mathbb{E}\big[\prod_{i=1}^{d} \lambda_i^{n_{\lambda_i}}\big], \quad \psi(\boldsymbol{n_Q}, \mathbf{0}) = \mathbb{E}\big[\prod_{i=1}^{d} Q_i^{[n_{Q_i}]}\big], \tag{75}$$

by straightforward substitution. Indeed, for fixed $n_\lambda \in \mathbb{N}$, we substitute $n_{Q_j} \equiv 0$ for all $j \in [d]$ in Eqn. (23). Rearranging terms, we obtain

$$\begin{aligned}\psi(\mathbf{0}, \boldsymbol{n_\lambda}) = \Big(\sum_{j=1}^{d} n_{\lambda_j}(\alpha_j - \mathbb{E}[B_{jj}])\Big)^{-1} \sum_{j=1}^{d} \Big\{ \sum_{\substack{i=1\\ i\neq j}}^{d} n_{\lambda_i} \mathbb{E}\big[B_{ij}\big] \psi(\mathbf{0}, \boldsymbol{n_\lambda} - \boldsymbol{e}_i + \boldsymbol{e}_j) \qquad (76)\\ + \alpha_j \overline{\lambda}_j n_{\lambda_j} \psi(\mathbf{0}, \boldsymbol{n_\lambda} - \boldsymbol{e}_j) + \sum_{m_1=0}^{n_{\lambda_1}} \cdots \sum_{m_d=0}^{n_{\lambda_d}} \mathbf{1}_{\{m \leqslant n_\lambda - 2\}} \prod_{k=1}^{d} \binom{n_{\lambda_k}}{m_k} \prod_{i=1}^{d} \mathbb{E}\big[B_{ij}^{n_{\lambda_i} - m_i}\big] \psi(\mathbf{0}, \boldsymbol{m} + \boldsymbol{e}_j) \Big\}.\end{aligned}$$

Observe that in order to obtain a final closed-form expression for $\psi(\mathbf{0}, \boldsymbol{n_\lambda})$, we need to solve a linear system of equations of equal order moments, i.e., the $\psi(\mathbf{0}, \boldsymbol{n_\lambda} - \boldsymbol{e}_i + \boldsymbol{e}_j)$ terms.

A similar result holds for the joint moments of $\boldsymbol{Q}$: for fixed $n_Q \in \mathbb{N}$, we substitute $n_{\lambda_j} \equiv 0$ for all $j \in [d]$ in Eqn. (23), yielding

$$\psi(\boldsymbol{n_Q}, \mathbf{0}) = \Big(\sum_{j=1}^{d} n_{Q_j} \mu_j\Big)^{-1} \sum_{j=1}^{d} n_{Q_j} \psi(\boldsymbol{n_Q} - \boldsymbol{e}_j, \boldsymbol{e_j}). \tag{77}$$

Appendix D. Explicit Examples for Bivariate Setting

In this section, we provide more explicit details for the moments in the bivariate setting, i.e, the case $d = 2$. We provide examples by writing out the recursive procedure outlined in Section 4. The main objective is to derive near-explicit results for both the transient moments $\psi_t((n_{Q_1}, n_{Q_2}), (n_{\lambda_1}, n_{\lambda_2}))$ and stationary moments $\psi((n_{Q_1}, n_{Q_2}), (n_{\lambda_1}, n_{\lambda_2}))$, where the focus is on moments of order 1 and 2. In both cases, we apply the recursive procedures described in Section 4.

D.1. **Recursive procedure.** We illustrate the stacked vector $\boldsymbol{\Psi}_t^{(n)}$ for orders $n = 1$ and $n = 2$, and derive the ODEs associated with the recursive procedure.

Example 1 (first order, bivariate)**.** For $n = 1$, we have $\mathfrak{D}(2, 1) = 4$, and

$$\boldsymbol{\Psi}_t^{(1)} = \big(\boldsymbol{\Psi}_t^{(0,1)}, \boldsymbol{\Psi}_t^{(1,0)}\big)^\top,$$

where

$$\boldsymbol{\Psi}_t^{(0,1)} = \big(\mathbb{E}\big[\lambda_1(t)\big], \mathbb{E}\big[\lambda_2(t)\big]\big)^\top, \quad \boldsymbol{\Psi}_t^{(1,0)} = \big(\mathbb{E}\big[Q_1(t)\big], \mathbb{E}\big[Q_2(t)\big]\big)^\top.$$

By Step 0 of Algorithm 1, we obtain the ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{\Psi}_t^{(0,1)} = \begin{bmatrix} -\overline{\alpha}_1 & \mathbb{E}[B_{12}] \\ \mathbb{E}[B_{21}] & -\overline{\alpha}_2 \end{bmatrix} \mathbf{\Psi}_t^{(0,1)} + \begin{bmatrix} \alpha_1\overline{\lambda}_1 \\ \alpha_2\overline{\lambda}_2 \end{bmatrix}, \tag{78}$$

whose solution gives us an expression for $\mathbf{\Psi}_t^{(0,1)} = (\mathbb{E}[\lambda_1(t)], \mathbb{E}[\lambda_2(t)])^\top$. We need this $\mathbf{\Psi}_t^{(0,1)}$ in Step 1, which states

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{\Psi}_t^{(1,0)} = \begin{bmatrix} -\mu_1 & 0 \\ 0 & -\mu_2 \end{bmatrix} \mathbf{\Psi}_t^{(1,0)} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{\Psi}_t^{(0,1)}, \tag{79}$$

whose solution yields an expression for $\mathbf{\Psi}_t^{(1,0)} = (\mathbb{E}[Q_1(t)], \mathbb{E}[Q_2(t)])^\top$.

Example 2 (second order, bivariate)**.** For $n = 2$, we have $\mathfrak{D}(2,2) = 10$, and

$$\mathbf{\Psi}_t^{(2)} = \left(\mathbf{\Psi}_t^{(0,2)}, \mathbf{\Psi}_t^{(1,1)}, \mathbf{\Psi}_t^{(2,0)}\right)^\top,$$

where

$$\begin{aligned}
\mathbf{\Psi}_t^{(0,2)} &= \left(\mathbb{E}\big[\lambda_1(t)^2\big], \mathbb{E}\big[\lambda_1(t)\lambda_2(t)\big], \mathbb{E}\big[\lambda_2(t)^2\big]\right)^\top, \\
\mathbf{\Psi}_t^{(1,1)} &= \left(\mathbb{E}\big[Q_1(t)\lambda_1(t)\big], \mathbb{E}\big[Q_1(t)\lambda_2(t)\big], \mathbb{E}\big[Q_2(t)\lambda_1(t)\big], \mathbb{E}\big[Q_2(t)\lambda_2(t)\big]\right)^\top, \\
\mathbf{\Psi}_t^{(2,0)} &= \left(\mathbb{E}\big[Q_1(t)^{[2]}\big], \mathbb{E}\big[Q_1(t)Q_2(t)\big], \mathbb{E}\big[Q_2(t)^{[2]}\big]\right)^\top.
\end{aligned}$$

For order $n = 2$, our objective is to compute $\mathbf{\Psi}_t^{(0,2)}$, $\mathbf{\Psi}_t^{(1,1)}$, and $\mathbf{\Psi}_t^{(2,0)}$. Step 0 of Algorithm 1 yields

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{\Psi}_t^{(0,2)} &= \begin{bmatrix} -2\overline{\alpha}_1 & 2\mathbb{E}[B_{12}] & 0 \\ \mathbb{E}[B_{21}] & -\overline{\alpha}_1 - \overline{\alpha}_2 & \mathbb{E}[B_{12}] \\ 0 & 2\mathbb{E}[B_{21}] & -2\overline{\alpha}_2 \end{bmatrix} \mathbf{\Psi}_t^{(0,2)} \\
&\quad + \begin{bmatrix} 2\alpha_1\overline{\lambda}_1 + \mathbb{E}[B_{11}^2] & \mathbb{E}[B_{12}^2] \\ \mathbb{E}[B_{11}]\mathbb{E}[B_{21}] + \alpha_2\overline{\lambda}_2 & \mathbb{E}[B_{22}]\mathbb{E}[B_{12}] + \alpha_1\overline{\lambda}_1 \\ \mathbb{E}[B_{21}^2] & 2\alpha_2\overline{\lambda}_2 + \mathbb{E}[B_{22}^2] \end{bmatrix} \mathbf{\Psi}_t^{(0,1)},
\end{aligned} \tag{80}$$

which depends on the lower-order vector $\mathbf{\Psi}_t^{(0,1)}$ (which was found in Example 1). For Step 1, note that $\mathbf{\Psi}_t^{(1,1)}$ is a 4-dimensional vector, which satisfies

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{\Psi}_t^{(1,1)} &= \begin{bmatrix} -\overline{\alpha}_1 - \mu_1 & \mathbb{E}[B_{12}] \\ \mathbb{E}[B_{21}] & -\overline{\alpha}_2 - \mu_1 \end{bmatrix} \oplus \begin{bmatrix} -\overline{\alpha}_1 - \mu_2 & \mathbb{E}[B_{12}] \\ \mathbb{E}[B_{21}] & -\overline{\alpha}_2 - \mu_2 \end{bmatrix} \mathbf{\Psi}_t^{(1,1)} \\
&\quad + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{\Psi}_t^{(0,2)} + \begin{bmatrix} \alpha_1\overline{\lambda}_1 & 0 & \mathbb{E}[B_{11}] & 0 \\ \alpha_2\overline{\lambda}_2 & 0 & \mathbb{E}[B_{21}] & 0 \\ 0 & \alpha_1\overline{\lambda}_1 & 0 & \mathbb{E}[B_{12}] \\ 0 & \alpha_2\overline{\lambda}_2 & 0 & \mathbb{E}[B_{22}] \end{bmatrix} \mathbf{\Psi}_t^{(1)},
\end{aligned} \tag{81}$$

where we see the dependence on the lower-order stacked vector $\boldsymbol{\Psi}_t^{(1)}$ (which was found in Example 1). Regarding the final step, i.e., Step 2,

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Psi}_t^{(2,0)} = \begin{bmatrix} -2\mu_1 & 0 & 0 \\ 0 & -\mu_1-\mu_2 & 0 \\ 0 & 0 & -2\mu_2 \end{bmatrix} \boldsymbol{\Psi}_t^{(2,0)} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \boldsymbol{\Psi}_t^{(1,1)}. \tag{82}$$

Example 3 (third order, bivariate)**.** For $n = 3$, we have $\mathfrak{D}(2,3) = 20$, and

$$\boldsymbol{\Psi}_t^{(3)} = \big(\boldsymbol{\Psi}_t^{(0,3)}, \boldsymbol{\Psi}_t^{(1,2)}, \boldsymbol{\Psi}_t^{(2,1)}, \boldsymbol{\Psi}_t^{(3,0)}\big)^\top,$$

where

$$\begin{aligned}
\boldsymbol{\Psi}_t^{(0,3)} &= \big(\mathbb{E}\big[\lambda_1(t)^3\big], \mathbb{E}\big[\lambda_1(t)^2\lambda_2(t)\big], \mathbb{E}\big[\lambda_1(t)\lambda_2(t)^2\big], \mathbb{E}\big[\lambda_2(t)^3\big]\big)^\top, \\
\boldsymbol{\Psi}_t^{(1,2)} &= \big(\mathbb{E}\big[Q_1(t)\lambda_1(t)^2\big], \mathbb{E}\big[Q_1(t)\lambda_1\lambda_2(t)\big], \mathbb{E}\big[Q_1(t)\lambda_2(t)^2\big], \\
&\qquad \mathbb{E}\big[Q_2(t)\lambda_1(t)^2\big], \mathbb{E}\big[Q_2(t)\lambda_1(t)\lambda_2\big], \mathbb{E}\big[Q_2(t)\lambda_2(t)^2\big]\big)^\top, \\
\boldsymbol{\Psi}_t^{(2,1)} &= \big(\mathbb{E}\big[Q_1(t)^{[2]}\lambda_1(t)\big], \mathbb{E}\big[Q_1(t)^{[2]}\lambda_2(t)\big], \mathbb{E}\big[Q_1(t)Q_2(t)\lambda_1(t)\big], \mathbb{E}\big[Q_1(t)Q_2(t)\lambda_2(t)\big], \\
&\qquad \mathbb{E}\big[Q_2(t)^{[2]}\lambda_1(t)\big], \mathbb{E}\big[Q_2(t)^{[2]}\lambda_2(t)\big]\big)^\top, \\
\boldsymbol{\Psi}_t^{(3,0)} &= \big(\mathbb{E}\big[Q_1(t)^{[3]}\big], \mathbb{E}\big[Q_1(t)^{[2]}Q_2(t)\big], \mathbb{E}\big[Q_1(t)Q_2(t)^{[2]}\big], \mathbb{E}\big[Q_2(t)^{[3]}\big]\big)^\top.
\end{aligned}$$

We could explicitly write down the ODEs of these vectors using Algorithm 1, but the exposition would be rather tedious with large matrices.

D.2. **Transient moments.** The goal of this subsection is to find near-explicit expressions for $\boldsymbol{\Psi}_t^{(1)}$ and $\boldsymbol{\Psi}_t^{(2)}$ by further solving the associated ODEs. It is clear that we can obtain the solution in terms of a matrix exponential, which can be made more explicit in terms of its eigenvalues, namely

$$e^{t\boldsymbol{M}^{(k,n-k)}} = \sum_{\ell=1}^{\overline{k}} e^{t\eta_\ell^{(k)}} \prod_{\substack{m=1\\ m\neq\ell}}^{\overline{k}} \frac{\boldsymbol{M}^{(k,n-k)} - \eta_m^{(k)}\boldsymbol{I}}{\eta_\ell^{(k)} - \eta_m^{(k)}}, \tag{83}$$

with $\overline{k}$ denoting the dimension of $\boldsymbol{\Psi}_t^{(k,n-k)}$ and $\eta_1^{(k)}, \dots, \eta_{\overline{k}}^{(k)}$ the eigenvalues of $\boldsymbol{M}^{(k,n-k)}$, and $\boldsymbol{I}$ the identity matrix.

We first consider the transient moments of order $n = 1$, evaluating the entries of the stacked vector $\boldsymbol{\Psi}_t^{(1)}$, in particular solving the ODE of $\boldsymbol{\Psi}_t^{(0,1)}$ as given in Eqn. (78). By Proposition 1, the solution requires us to find the eigenvalues of the matrix

$$\boldsymbol{M}^{(0,1)} = \begin{bmatrix} -\overline{\alpha}_1 & \mathbb{E}[B_{12}] \\ \mathbb{E}[B_{21}] & -\overline{\alpha}_2 \end{bmatrix},$$

so as to compute the matrix exponential $e^{t\boldsymbol{M}^{(0,1)}}$. With $\eta \equiv \eta_1, \eta_2$ denoting the two eigenvalues, we straightforwardly obtain

$$\eta = \frac{1}{2}\big(-\overline{\alpha}_1 - \overline{\alpha}_2 \pm \sqrt{\overline{\alpha}_1^2 - 2\overline{\alpha}_1\overline{\alpha}_2 + \overline{\alpha}_2^2 + 4\mathbb{E}[B_{12}]\mathbb{E}[B_{21}]}\big) \equiv \frac{1}{2}\big(-\overline{\alpha}_1 - \overline{\alpha}_2 \pm \sqrt{D_1}\big), \tag{84}$$

where $\overline{\alpha}_i = \alpha_i - \mathbb{E}[B_{ii}]$ for $i = 1, 2$. We let η_1 and η_2 denote the plus- and minus-variant of η respectively. Note that $D_1 \geqslant 0$ since it involves a square and B_{ij} are non-negative random variables. Using Eqn. (83) and performing some elementary computations,

$$\begin{aligned} e^{t\boldsymbol{M}^{(0,1)}} &= \frac{1}{\eta_1 - \eta_2}\Big(e^{t\eta_1}\big(\boldsymbol{M}^{(0,1)} - \eta_2 \boldsymbol{I}\big) - e^{t\eta_2}\big(\boldsymbol{M}^{(0,1)} - \eta_1 \boldsymbol{I}\big)\Big) \\ &= \frac{1}{\sqrt{D_1}}\Big(\big\{e^{t\eta_1} - e^{t\eta_2}\big\}\begin{bmatrix} \frac{1}{2}(\overline{\alpha}_2 - \overline{\alpha}_1) & \mathbb{E}[B_{12}] \\ \mathbb{E}[B_{21}] & \frac{1}{2}(\overline{\alpha}_1 - \overline{\alpha}_2) \end{bmatrix} + \big\{e^{t\eta_1} + e^{t\eta_2}\big\}\begin{bmatrix} \frac{1}{2}\sqrt{D_1} & 0 \\ 0 & \frac{1}{2}\sqrt{D_1} \end{bmatrix}\Big) \\ &= \frac{1}{2}\big\{e^{t\eta_1} + e^{t\eta_2}\big\}\boldsymbol{I} + \frac{1}{\sqrt{D_1}}\big\{e^{t\eta_1} - e^{t\eta_2}\big\}\begin{bmatrix} \frac{1}{2}(\overline{\alpha}_2 - \overline{\alpha}_1) & \mathbb{E}[B_{12}] \\ \mathbb{E}[B_{21}] & \frac{1}{2}(\overline{\alpha}_1 - \overline{\alpha}_2) \end{bmatrix}; \end{aligned}$$

we use curly brackets to distinguish scalar terms from the vectors and matrices.

Next, we consider $\boldsymbol{\Psi}_t^{(1,0)}$ for which we need to find the eigenvalues of the matrix $\boldsymbol{M}^{(1,0)} = \operatorname{diag}(-\mu_1, -\mu_2)$; cf. the ODE in Eqn. (79). Since this is a diagonal matrix, these are simply $-\mu_1$ and $-\mu_2$, such that the matrix exponential is just

$$e^{t\boldsymbol{M}^{(1,0)}} = \begin{bmatrix} e^{-t\mu_1} & 0 \\ 0 & e^{-t\mu_2} \end{bmatrix}.$$

Before we get to the solution, we define a number of functions needed for the solution of $\boldsymbol{\Psi}_t^{(1,0)}$, namely

$$\begin{aligned} \boldsymbol{u}_1(t) &:= \begin{bmatrix} (\mu_1 + \eta_1)^{-1}(e^{t\eta_1} - e^{-t\mu_1}) + (\mu_1 + \eta_2)^{-1}(e^{t\eta_2} - e^{-t\mu_1}) \\ (\mu_2 + \eta_1)^{-1}(e^{t\eta_1} - e^{-t\mu_2}) + (\mu_2 + \eta_2)^{-1}(e^{t\eta_2} - e^{-t\mu_2}) \end{bmatrix}, \\ \boldsymbol{u}_2(t) &:= \begin{bmatrix} (\mu_1 + \eta_1)^{-1}(e^{t\eta_1} - e^{-t\mu_1}) - (\mu_1 + \eta_2)^{-1}(e^{t\eta_2} - e^{-t\mu_1}) \\ (\mu_2 + \eta_1)^{-1}(e^{t\eta_1} - e^{-t\mu_2}) - (\mu_2 + \eta_2)^{-1}(e^{t\eta_2} - e^{-t\mu_2}) \end{bmatrix}, \\ \boldsymbol{u}_3(t) &:= \begin{bmatrix} (\eta_1(\mu_1 + \eta_1))^{-1}(e^{t\eta_1} - e^{-t\mu_1}) + (\eta_2(\mu_1 + \eta_2))^{-1}(e^{t\eta_2} - e^{-t\mu_1}) \\ (\eta_1(\mu_2 + \eta_1))^{-1}(e^{t\eta_1} - e^{-t\mu_2}) + (\eta_2(\mu_2 + \eta_2))^{-1}(e^{t\eta_2} - e^{-t\mu_2}) \end{bmatrix}, \\ \boldsymbol{u}_4(t) &:= \begin{bmatrix} (\eta_1(\mu_1 + \eta_1))^{-1}(e^{t\eta_1} - e^{-t\mu_1}) - (\eta_2(\mu_1 + \eta_2))^{-1}(e^{t\eta_2} - e^{-t\mu_1}) \\ (\eta_1(\mu_2 + \eta_1))^{-1}(e^{t\eta_1} - e^{-t\mu_2}) - (\eta_2(\mu_2 + \eta_2))^{-1}(e^{t\eta_2} - e^{-t\mu_2}) \end{bmatrix}. \end{aligned}$$

We can now give the first moments explicitly through the application of Proposition 1. Given the initial conditions $\boldsymbol{\Psi}_0^{(0,1)} = (\overline{\lambda}_1, \overline{\lambda}_2)^\top$ and $\boldsymbol{\Psi}_0^{(1,0)} = (0, 0)^\top$, the solutions to the ODEs in (78) and (79) are given by

$$
\begin{aligned}
\boldsymbol{\Psi}_t^{(0,1)} &= e^{t\boldsymbol{M}^{(0,1)}}\begin{bmatrix}\overline{\lambda}_1\\ \overline{\lambda}_2\end{bmatrix} + \int_0^t e^{(t-s)\boldsymbol{M}^{(0,1)}}\begin{bmatrix}\alpha_1\overline{\lambda}_1\\ \alpha_2\overline{\lambda}_2\end{bmatrix}\mathrm{d}s\\
&= \frac{1}{\overline{\alpha}_1\overline{\alpha}_2-\mathbb{E}[B_{12}]\mathbb{E}[B_{21}]}\begin{bmatrix}\alpha_1\overline{\lambda}_1\overline{\alpha}_2+\alpha_2\overline{\lambda}_2\mathbb{E}[B_{12}]\\ \alpha_2\overline{\lambda}_2\overline{\alpha}_1+\alpha_1\overline{\lambda}_1\mathbb{E}[B_{21}]\end{bmatrix}\\
&\quad+\frac{1}{2}\left\{e^{t\eta_1}+e^{t\eta_2}\right\}\begin{bmatrix}\overline{\lambda}_1\\ \overline{\lambda}_2\end{bmatrix}+\frac{1}{\sqrt{D_1}}\left\{e^{t\eta_1}-e^{t\eta_2}\right\}\begin{bmatrix}\frac{1}{2}\overline{\lambda}_1(\overline{\alpha}_2-\overline{\alpha}_1)+\overline{\lambda}_2\mathbb{E}[B_{12}]\\ \frac{1}{2}\overline{\lambda}_2(\overline{\alpha}_1-\overline{\alpha}_2)+\overline{\lambda}_1\mathbb{E}[B_{21}]\end{bmatrix}\\
&\quad+\frac{1}{2}\left\{\eta_1^{-1}e^{t\eta_1}+\eta_2^{-1}e^{t\eta_2}\right\}\begin{bmatrix}\alpha_1\overline{\lambda}_1\\ \alpha_2\overline{\lambda}_2\end{bmatrix}\\
&\quad+\frac{1}{\sqrt{D_1}}\left\{\eta_1^{-1}e^{t\eta_1}-\eta_2^{-1}e^{t\eta_2}\right\}\begin{bmatrix}\frac{1}{2}\alpha_1\overline{\lambda}_1(\overline{\alpha}_2-\overline{\alpha}_1)+\alpha_2\overline{\lambda}_2\mathbb{E}[B_{12}]\\ \frac{1}{2}\alpha_2\overline{\lambda}_2(\overline{\alpha}_1-\overline{\alpha}_2)+\alpha_1\overline{\lambda}_1\mathbb{E}[B_{21}]\end{bmatrix},
\end{aligned}
\tag{85}
$$

and

$$
\begin{aligned}
\boldsymbol{\Psi}_t^{(1,0)} &= \int_0^t e^{(t-s)\boldsymbol{M}^{(1,0)}}\boldsymbol{\Psi}_s^{(0,1)}\mathrm{d}s\\
&= \frac{1}{\overline{\alpha}_1\overline{\alpha}_2-\mathbb{E}[B_{12}]\mathbb{E}[B_{21}]}\begin{bmatrix}\mu_1^{-1}(1-e^{-t\mu_1})\\ \mu_1^{-2}(1-e^{-t\mu_2})\end{bmatrix}\odot\begin{bmatrix}\alpha_1\overline{\lambda}_1\overline{\alpha}_2+\alpha_2\overline{\lambda}_2\mathbb{E}[B_{12}]\\ \alpha_2\overline{\lambda}_2\overline{\alpha}_1+\alpha_1\overline{\lambda}_1\mathbb{E}[B_{21}]\end{bmatrix}\\
&\quad+\frac{1}{2}\boldsymbol{u}_1(t)\odot\begin{bmatrix}\overline{\lambda}_1\\ \overline{\lambda}_2\end{bmatrix}+\frac{1}{\sqrt{D_1}}\boldsymbol{u}_2(t)\odot\begin{bmatrix}\frac{1}{2}\overline{\lambda}_1(\overline{\alpha}_2-\overline{\alpha}_1)+\overline{\lambda}_2\mathbb{E}[B_{12}]\\ \frac{1}{2}\overline{\lambda}_2(\overline{\alpha}_1-\overline{\alpha}_2)+\overline{\lambda}_1\mathbb{E}[B_{21}]\end{bmatrix}\\
&\quad+\frac{1}{2}\boldsymbol{u}_3(t)\odot\begin{bmatrix}\alpha_1\overline{\lambda}_1\\ \alpha_2\overline{\lambda}_2\end{bmatrix}+\frac{1}{\sqrt{D_1}}\boldsymbol{u}_4(t)\odot\begin{bmatrix}\frac{1}{2}\alpha_1\overline{\lambda}_1(\overline{\alpha}_2-\overline{\alpha}_1)+\alpha_2\overline{\lambda}_2\mathbb{E}[B_{12}]\\ \frac{1}{2}\alpha_2\overline{\lambda}_2(\overline{\alpha}_1-\overline{\alpha}_2)+\alpha_1\overline{\lambda}_1\mathbb{E}[B_{21}]\end{bmatrix}.
\end{aligned}
\tag{86}
$$

Observe that in order for the solution in (85) to remain stable and to obtain finite moments, we need that both eigenvalues are strictly smaller than 0. By some elementary algebra, it is seen that we should have that

$$\overline{\alpha}_1\overline{\alpha}_2 > \mathbb{E}[B_{12}]\mathbb{E}[B_{21}]. \tag{87}$$

Note that this is the explicit version of the stability condition $\rho(\boldsymbol{H}) < 1$ for the bivariate setting; see Assumption 1. Also note that if one is interested in the Hawkes process $\boldsymbol{N}(t) = (N_1(t), N_2(t))^\top$ rather than the population process $\boldsymbol{Q}(t) = (Q_1(t), Q_2(t))^\top$, one needs to take $\mu_1 = \mu_2 \equiv 0$. The corresponding moments $\mathbb{E}[\boldsymbol{N}(t)] = (\mathbb{E}[N_1(t)], \mathbb{E}[N_2(t)])^\top$ can be derived from Eqn. (86) by taking the limit $(\mu_1, \mu_2) \downarrow (0, 0)$ and the use of L'Hopital's rule.

We now turn to order 2 and compute elements of the stacked vector $\boldsymbol{\Psi}_t^{(2)}$. As before, we start by considering $\boldsymbol{\Psi}_t^{(0,2)}$, the vector containing the (mixed) moments corresponding to $\boldsymbol{\lambda}(t)$. By Proposition 1, and recalling that $\boldsymbol{\Psi}_t^{(0,2)}$ satisfies the ODE in Eqn. (80), we need to find the eigenvalues of

$$\boldsymbol{M}^{(0,2)} = \begin{bmatrix}-2\overline{\alpha}_1 & 2\mathbb{E}[B_{12}] & 0\\ \mathbb{E}[B_{21}] & -\overline{\alpha}_1-\overline{\alpha}_2 & \mathbb{E}[B_{12}]\\ 0 & 2\mathbb{E}[B_{21}] & -2\overline{\alpha}_2\end{bmatrix}.$$

Let $\kappa \equiv \kappa_1, \kappa_2, \kappa_3$ denote the eigenvalues of $\boldsymbol{M}^{(0,2)}$. We compute

$$\begin{aligned}
&\begin{vmatrix} -2\overline{\alpha}_1-\kappa & 2\mathbb{E}[B_{12}] & 0 \\ \mathbb{E}[B_{21}] & -\overline{\alpha}_1-\overline{\alpha}_2-\kappa & \mathbb{E}[B_{12}] \\ 0 & 2\mathbb{E}[B_{21}] & -2\overline{\alpha}_2-\kappa \end{vmatrix} = 0 \\
\iff &(-2\overline{\alpha}_1-\kappa)\begin{vmatrix} -\overline{\alpha}_1-\overline{\alpha}_2-\kappa & \mathbb{E}[B_{12}] \\ 2\mathbb{E}[B_{21}] & -2\overline{\alpha}_2-\kappa \end{vmatrix} - 2\mathbb{E}[B_{12}]\begin{vmatrix} \mathbb{E}[B_{21}] & \mathbb{E}[B_{12}] \\ 0 & -2\overline{\alpha}_2-\kappa \end{vmatrix} = 0 \\
\iff &(-2\overline{\alpha}_1-\kappa)\big(\kappa^2+3\overline{\alpha}_2\kappa+2\overline{\alpha}_2^2+2\overline{\alpha}_1\overline{\alpha}_2+\overline{\alpha}_1\kappa-2\mathbb{E}[B_{12}]\mathbb{E}[B_{21}]\big) \\
&\quad + 2\mathbb{E}[B_{12}](\mathbb{E}[B_{21}]\kappa+2\overline{\alpha}_2\mathbb{E}[B_{21}] = 0 \\
\iff &\kappa^3+\kappa^2 3(\overline{\alpha}_1+\overline{\alpha}_2)+\kappa\big(2(\overline{\alpha}_1+\overline{\alpha}_2)^2-4\mathbb{E}[B_{12}]\mathbb{E}[B_{21}]\big)-4(\overline{\alpha}_1+\overline{\alpha}_2)\mathbb{E}[B_{12}]\mathbb{E}[B_{21}] = 0 \\
\iff &\kappa^3+\kappa^2 b+\kappa c+d = 0,
\end{aligned}$$

with b, c, d defined as the constants of the square, linear and constant term respectively. To apply the formula for the solutions to this cubic equation, we compute p and q, given by

$$\begin{aligned}
p = \frac{1}{3}(3c-b^2) &= 2(\overline{\alpha}_1+\overline{\alpha}_2)^2-4\mathbb{E}[B_{12}]\mathbb{E}[B_{21}]-3((\overline{\alpha}_1+\overline{\alpha}_2)^2 \\
&= -(\overline{\alpha}_1+\overline{\alpha}_2)^2-4\mathbb{E}[B_{12}]\mathbb{E}[B_{21}],
\end{aligned}$$

and

$$\begin{aligned}
q &= \frac{1}{27}\big\{2b^3-9bc+27d\big\} \\
&= \frac{1}{27}\Big\{2\big(3(\overline{\alpha}_1+\overline{\alpha}_2)\big)^3-27(\overline{\alpha}_1+\overline{\alpha}_2)\big(2(\overline{\alpha}_1+\overline{\alpha}_2)^2-4\mathbb{E}[B_{12}]\mathbb{E}[B_{21}]\big) \\
&\qquad -4\cdot 27(\overline{\alpha}_1+\overline{\alpha}_2)\mathbb{E}[B_{12}]\mathbb{E}[B_{21}]\Big\} \\
&= 2(\overline{\alpha}_1+\overline{\alpha}_2)^3-2(\overline{\alpha}_1+\overline{\alpha}_2)^3+4(\overline{\alpha}_1+\overline{\alpha}_2)\mathbb{E}[B_{12}]\mathbb{E}[B_{21}]-4(\overline{\alpha}_1+\overline{\alpha}_2)\mathbb{E}[B_{12}]\mathbb{E}[B_{21}] \\
&= 0.
\end{aligned}$$

It is well-known that the cubic equation has three real roots if $4p^3+27q^2<0$. Since $q=0$, the condition becomes $4p^3<0$, which holds since $p<0$ because of the square term and $\mathbb{E}[B_{ij}] \geqslant 0$. Hence, the eigenvalues κ_m, with $m=1,2,3$, are given by the trigonometric solution

$$\begin{aligned}
\kappa_k &= -\frac{b}{3}+2\sqrt{\frac{-p}{3}}\cos(\theta_m) = -(\overline{\alpha}_1+\overline{\alpha}_2)+2\sqrt{(\overline{\alpha}_1+\overline{\alpha}_2)^2+4\mathbb{E}[B_{12}]\mathbb{E}[B_{21}]}\cos(\theta_m), \\
\theta_m &= \frac{1}{3}\arccos\Big(\frac{3q}{2p}\sqrt{\frac{3}{-p}}\Big)-\frac{2\pi}{3}(m-1) = \frac{\pi}{6}-\frac{2\pi}{3}(m-1).
\end{aligned}$$

This yields the eigenvalues

$$\begin{aligned}
\kappa_1 &= -(\overline{\alpha}_1+\overline{\alpha}_2) \\
\kappa_2 &= -(\overline{\alpha}_1+\overline{\alpha}_2)+\sqrt{3}\sqrt{(\overline{\alpha}_1+\overline{\alpha}_2)^2+4\mathbb{E}[B_{12}]\mathbb{E}[B_{21}]} = -(\overline{\alpha}_1+\overline{\alpha}_2)+\sqrt{D_2} \\
\kappa_3 &= -(\overline{\alpha}_1+\overline{\alpha}_2)-\sqrt{3}\sqrt{(\overline{\alpha}_1+\overline{\alpha}_2)^2+4\mathbb{E}[B_{12}]\mathbb{E}[B_{21}]} = -(\overline{\alpha}_1+\overline{\alpha}_2)-\sqrt{D_2},
\end{aligned}$$

with $D_2 = 3((\overline{\alpha}_1+\overline{\alpha}_2)^2 + 4\mathbb{E}[B_{12}]\mathbb{E}[B_{21}])$. We apply these eigenvalues in the computation of the matrix exponential, as described in Eqn. (83), to obtain

$$
\begin{aligned}
e^{t\boldsymbol{M}^{(0,2)}} &= e^{\kappa_1 t}\frac{1}{\kappa_1-\kappa_2}\frac{1}{\kappa_1-\kappa_3}(\boldsymbol{M}^{(0,2)}-\kappa_2\boldsymbol{I})(\boldsymbol{M}^{(0,2)}-\kappa_3\boldsymbol{I})\\
&\quad + e^{\kappa_2 t}\frac{1}{\kappa_2-\kappa_1}\frac{1}{\kappa_2-\kappa_3}(\boldsymbol{M}^{(0,2)}-\kappa_1\boldsymbol{I})(\boldsymbol{M}^{(0,2)}-\kappa_3\boldsymbol{I})\\
&\quad + e^{\kappa_3 t}\frac{1}{\kappa_3-\kappa_1}\frac{1}{\kappa_3-\kappa_2}(\boldsymbol{M}^{(0,2)}-\kappa_1\boldsymbol{I})(\boldsymbol{M}^{(0,2)}-\kappa_2\boldsymbol{I})\\
&= \frac{e^{\kappa_1 t}}{D_2}\begin{bmatrix} c_{\kappa_1} & -2\mathbb{E}[B_{12}](\overline{\alpha}_2-\overline{\alpha}_1) & -2\mathbb{E}[B_{12}]^2\\ -\mathbb{E}[B_{21}](\overline{\alpha}_2-\overline{\alpha}_1) & 3(\overline{\alpha}_1+\overline{\alpha}_2)^2+8\mathbb{E}[B_{12}]\mathbb{E}[B_{21}] & -\mathbb{E}[B_{12}](\overline{\alpha}_1-\overline{\alpha}_2)\\ -2\mathbb{E}[B_{21}]^2 & -2\mathbb{E}[B_{21}](\overline{\alpha}_1-\overline{\alpha}_2) & c_{\kappa_1}\end{bmatrix}\\
&\quad + \frac{e^{\kappa_2 t}}{2D_2}\begin{bmatrix} c_- & -2\mathbb{E}[B_{12}](\overline{\alpha}_1-\overline{\alpha}_2-\sqrt{D_2}) & 2\mathbb{E}[B_{12}]^2\\ -\mathbb{E}[B_{21}](\overline{\alpha}_1-\overline{\alpha}_2-\sqrt{D_2}) & 4\mathbb{E}[B_{12}]\mathbb{E}[B_{21}] & \mathbb{E}[B_{12}](\overline{\alpha}_1-\overline{\alpha}_2+\sqrt{D_2})\\ 2\mathbb{E}[B_{21}]^2 & 2\mathbb{E}[B_{21}](\overline{\alpha}_1-\overline{\alpha}_2+\sqrt{D_2}) & c_+\end{bmatrix}\\
&\quad + \frac{e^{\kappa_3 t}}{2D_2}\begin{bmatrix} c_+ & -2\mathbb{E}[B_{12}](\overline{\alpha}_1-\overline{\alpha}_2+\sqrt{D_2}) & 2\mathbb{E}[B_{12}]^2\\ -\mathbb{E}[B_{21}](\overline{\alpha}_1-\overline{\alpha}_2-\sqrt{D_2}) & 4\mathbb{E}[B_{12}]\mathbb{E}[B_{21}] & \mathbb{E}[B_{12}](\overline{\alpha}_1-\overline{\alpha}_2-\sqrt{D_2})\\ 2\mathbb{E}[B_{21}]^2 & 2\mathbb{E}[B_{21}](\overline{\alpha}_1-\overline{\alpha}_2-\sqrt{D_2}) & c_-\end{bmatrix},
\end{aligned}
$$

where

$$
\begin{aligned}
c_{\kappa_1} &= 2\overline{\alpha}_1^2 + 8\overline{\alpha}_1\overline{\alpha}_2 + 2\overline{\alpha}_2^2 + 10\mathbb{E}[B_{12}]\mathbb{E}[B_{21}],\\
c_- &= 2\mathbb{E}[B_{12}]\mathbb{E}[B_{21}] + (\overline{\alpha}_1-\overline{\alpha}_2)(\overline{\alpha}_1-\overline{\alpha}_2-\sqrt{D_2}),\\
c_+ &= 2\mathbb{E}[B_{12}]\mathbb{E}[B_{21}] + (\overline{\alpha}_1-\overline{\alpha}_2)(\overline{\alpha}_1-\overline{\alpha}_2+\sqrt{D_2}).
\end{aligned}
$$

We now derive the matrix exponential corresponding to $\boldsymbol{\Psi}_t^{(1,1)}$, appearing in the ODE given in Eqn. (81). Observe that (81) reveals that to compute $\boldsymbol{\Psi}_t^{(1,1)}$, we need to know the non-homogeneous part of the equation, i.e. $\boldsymbol{\Psi}_t^{(0,2)}$, as well as the lower order ($n=1$) stacked vector $\boldsymbol{\Psi}_t^{(1)}$. Further notice that the 4×4 matrix $\boldsymbol{M}^{(1,1)}$ is the direct sum of two 2×2 matrices, which implies that we can split the 4-dimensional ODE into two 2-dimensional ODEs. We introduce the relevant objects by setting

$$\boldsymbol{\Psi}_t^{(1,1)} = \big(\boldsymbol{\Psi}_{t,Q_1}^{(1,1)}, \boldsymbol{\Psi}_{t,Q_2}^{(1,1)}\big)^\top,$$

where

$$
\begin{aligned}
\boldsymbol{\Psi}_{t,Q_1}^{(1,1)} &= \big(\mathbb{E}\big[Q_1(t)\lambda_1(t)\big], \mathbb{E}\big[Q_1(t)\lambda_2(t)\big]\big)^\top,\\
\boldsymbol{\Psi}_{t,Q_2}^{(1,1)} &= \big(\mathbb{E}\big[Q_2(t)\lambda_1(t)\big], \mathbb{E}\big[Q_2(t)\lambda_2(t)\big]\big)^\top.
\end{aligned}
$$

Focus on the solution of $\boldsymbol{\Psi}_{t,Q_1}^{(1,1)}$, where we note that the solution for $\boldsymbol{\Psi}_{t,Q_2}^{(1,1)}$ can be obtained in an analogous manner. Observe that one can derive from Eqn. (81) that $\boldsymbol{\Psi}_{t,Q_1}^{(1,1)}$ satisfies the ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Psi}_{t,Q_1}^{(1,1)} = \begin{bmatrix} -\overline{\alpha}_1-\mu_1 & \mathbb{E}[B_{12}]\\ \mathbb{E}[B_{21}] & -\overline{\alpha}_2-\mu_1\end{bmatrix}\boldsymbol{\Psi}_{t,Q_1}^{(1,1)} + \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}\boldsymbol{\Psi}_t^{(0,2)} + \begin{bmatrix}\alpha_1\overline{\lambda}_1 & 0 & \mathbb{E}[B_{11}] & 0\\ \alpha_2\overline{\lambda}_2 & 0 & \mathbb{E}[B_{21}] & 0\end{bmatrix}\boldsymbol{\Psi}_t^{(1)}.$$

This means that we need the matrix exponential of

$$\boldsymbol{M}_{Q_1}^{(1,1)} = \begin{bmatrix} -\overline{\alpha}_1 - \mu_1 & \mathbb{E}[B_{12}] \\ \mathbb{E}[B_{21}] & -\overline{\alpha}_2 - \mu_1 \end{bmatrix},$$

which requires us to find the corresponding two eigenvalues, denoted by $\gamma_1^{(Q_1)}$ and $\gamma_2^{(Q_1)}$. These eigenvalues are similarly derived as in Eqn. (84), in this case given by

$$\gamma_1^{(Q_1)} = -\mu_1 + \eta_1, \quad \gamma_2^{(Q_1)} = -\mu_1 + \eta_2,$$

with $\eta_1 = \frac{1}{2}\big(-\overline{\alpha}_1 - \overline{\alpha}_2 + \sqrt{D_1}\big)$ and $\eta_2 = \frac{1}{2}\big(-\overline{\alpha}_1 - \overline{\alpha}_2 - \sqrt{D_1}\big)$. Substituting this in the Lagrange interpolation formula of Eqn. (83), we obtain

$$\begin{aligned} e^{t\boldsymbol{M}_{Q_1}^{(1,1)}} &= \frac{1}{\gamma_1^{(Q_1)} - \gamma_2^{(Q_1)}}\Big\{e^{t\gamma_1^{(Q_1)}}\big(\boldsymbol{M}_{Q_1}^{(1,1)} - \gamma_2^{(Q_1)}\boldsymbol{I}_2\big) - e^{t\gamma_2^{(Q_1)}}\big(\boldsymbol{M}_{Q_1}^{(1,1)} - \gamma_1^{(Q_1)}\boldsymbol{I}_2\big)\Big\} \\ &= \frac{1}{2}\big\{e^{t\gamma_1^{(Q_1)}} + e^{t\gamma_2^{(Q_1)}}\big\}\boldsymbol{I} + \frac{1}{\sqrt{D_1}}\big\{e^{t\gamma_1^{(Q_1)}} - e^{t\gamma_2^{(Q_1)}}\big\}\begin{bmatrix} \frac{1}{2}(\overline{\alpha}_2 - \overline{\alpha}_1) & \mathbb{E}[B_{12}] \\ \mathbb{E}[B_{21}] & \frac{1}{2}(\overline{\alpha}_1 - \overline{\alpha}_2) \end{bmatrix}. \end{aligned}$$

In a very similar manner, the matrix exponential needed to evaluate $\boldsymbol{\Psi}_{t,Q_2}^{(1,1)}$ can be obtained from the ODE in Eqn. (81), and is given by

$$\begin{aligned} e^{t\boldsymbol{M}_{Q_2}^{(1,1)}} &= \frac{1}{\gamma_1^{(Q_2)} - \gamma_2^{(Q_2)}}\Big\{e^{t\gamma_1^{(Q_2)}}\big(\boldsymbol{M}_{Q_2}^{(1,1)} - \gamma_2^{(Q_2)}\boldsymbol{I}_2\big) - e^{t\gamma_2^{(Q_2)}}\big(\boldsymbol{M}_{Q_2}^{(1,1)} - \gamma_1^{(Q_2)}\boldsymbol{I}_2\big)\Big\} \\ &= \frac{1}{2}\big\{e^{t\gamma_1^{(Q_2)}} + e^{t\gamma_2^{(Q_2)}}\big\}\boldsymbol{I} + \frac{1}{\sqrt{D_1}}\big\{e^{t\gamma_1^{(Q_2)}} - e^{t\gamma_2^{(Q_2)}}\big\}\begin{bmatrix} \frac{1}{2}(\overline{\alpha}_2 - \overline{\alpha}_1) & \mathbb{E}[B_{12}] \\ \mathbb{E}[B_{21}] & \frac{1}{2}(\overline{\alpha}_1 - \overline{\alpha}_2) \end{bmatrix}, \end{aligned}$$

with $\gamma_1^{(Q_2)} = -\mu_2 + \eta_1$ and $\gamma_2^{(Q_2)} = -\mu_2 + \eta_2$.

Finally, for the solution of $\boldsymbol{\Psi}_t^{(2,0)}$, the vector containing the mixed factorial moments of $\boldsymbol{Q}(t)$, the matrix exponential of $\boldsymbol{M}^{(2,0)} = \mathrm{diag}(-2\mu_1, -\mu_1 - \mu_2, -2\mu_2)$ is simply

$$e^{t\boldsymbol{M}^{(2,0)}} = \begin{bmatrix} e^{-2t\mu_1} & 0 & 0 \\ 0 & e^{-t(\mu_1+\mu_2)} & 0 \\ 0 & 0 & e^{-2t\mu_2} \end{bmatrix}.$$

Applying Proposition 1 to $\boldsymbol{\Psi}_t^{(0,2)}$, $\boldsymbol{\Psi}_t^{(1,1)}$ and $\boldsymbol{\Psi}_t^{(2,0)}$, we obtain the following result. Given the initial conditions $\boldsymbol{\Psi}_0^{(0,2)} = (\overline{\lambda}_1^2, \overline{\lambda}_1\overline{\lambda}_2, \overline{\lambda}_2^2)^\top$, $\boldsymbol{\Psi}_{0,Q_1}^{(1,1)} = (0,0)^\top$, $\boldsymbol{\Psi}_{0,Q_2}^{(1,1)} = (0,0)^\top$ and $\boldsymbol{\Psi}_0^{(2,0)} = (0,0,0)^\top$, the solutions to the ODEs in Eqns. (80), (81) and (82), are given by, respectively,

$$\begin{aligned} \boldsymbol{\Psi}_t^{(0,2)} &= e^{t\boldsymbol{M}^{(0,2)}}\boldsymbol{\Psi}_0^{(0,2)} \\ &+ \int_0^t e^{(t-s)\boldsymbol{M}^{(0,2)}}\begin{bmatrix} 2\alpha_1\overline{\lambda}_1 + \mathbb{E}[B_{11}^2] & \mathbb{E}[B_{12}^2] \\ \mathbb{E}[B_{11}]\mathbb{E}[B_{21}] + \alpha_2\overline{\lambda}_2 & \mathbb{E}[B_{22}]\mathbb{E}[B_{12}] + \alpha_1\overline{\lambda}_1 \\ \mathbb{E}[B_{21}^2] & 2\alpha_2\overline{\lambda}_2 + \mathbb{E}[B_{22}^2] \end{bmatrix}\boldsymbol{\Psi}_s^{(0,1)}\mathrm{d}s, \end{aligned} \tag{88}$$

and $\boldsymbol{\Psi}_t^{(1,1)} = (\boldsymbol{\Psi}_{t,Q_1}^{(1,1)}, \boldsymbol{\Psi}_{t,Q_2}^{(1,1)})^\top$, with

$$
\begin{aligned}
\boldsymbol{\Psi}_{t,Q_1}^{(1,1)} &= \int_0^t e^{(t-s)\boldsymbol{M}_{Q_1}^{(1,1)}} \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \boldsymbol{\Psi}_s^{(0,2)} + \begin{bmatrix} \alpha_1\overline{\lambda}_1 & 0 & \mathbb{E}[B_{11}] & 0 \\ \alpha_2\overline{\lambda}_2 & 0 & \mathbb{E}[B_{21}] & 0 \end{bmatrix} \boldsymbol{\Psi}_s^{(1)} \right\} \mathrm{d}s \\
\boldsymbol{\Psi}_{t,Q_2}^{(1,1)} &= \int_0^t e^{(t-s)\boldsymbol{M}_{Q_2}^{(1,1)}} \left\{ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{\Psi}_s^{(0,2)} + \begin{bmatrix} 0 & \alpha_1\overline{\lambda}_1 & 0 & \mathbb{E}[B_{12}] \\ 0 & \alpha_2\overline{\lambda}_2 & 0 & \mathbb{E}[B_{22}] \end{bmatrix} \boldsymbol{\Psi}_s^{(1)} \right\} \mathrm{d}s
\end{aligned}
\tag{89}
$$

and

$$
\begin{aligned}
\boldsymbol{\Psi}_t^{(2,0)} &= \int_0^t e^{(t-s)\boldsymbol{M}^{(2,0)}} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \boldsymbol{\Psi}_s^{(1,1)} \mathrm{d}s \\
&= \int_0^t \begin{bmatrix} 2e^{-2(t-s)\mu_1}\mathbb{E}[Q_1(s)\lambda_1(s)] \\ e^{-(t-s)(\mu_1+\mu_2)}\big(\mathbb{E}[Q_1(s)\lambda_2(s)] + \mathbb{E}[Q_2(s)\lambda_1(s)]\big) \\ 2e^{-2(t-s)\mu_2}\mathbb{E}[Q_2(s)\lambda_2(s)] \end{bmatrix} \mathrm{d}s.
\end{aligned}
\tag{90}
$$

We remark that explicit evaluation of the expressions for the second-order moments is tedious, but in principle possible. For instance, it is evident that substituting the matrix exponential of $\boldsymbol{M}^{(0,2)}$ and the lower-order solution $\boldsymbol{\Psi}_s^{(0,1)}$ will yield rather involved expressions. For higher order moments, these expressions become even more complex. However, the near-explicit solution in terms of matrix exponentials and lower order terms is useful for practical purposes. Due to the availability of fast and robust algorithms for the matrix exponential and the solution of ODEs, it is relatively straightforward to numerically compute higher-order moments.

D.3. **Stationary moments.** This subsection deals with the application of Algorithm 2 to evaluate the stationary moments $\boldsymbol{\Psi}^{(1)}$ and $\boldsymbol{\Psi}^{(2)}$.

We start by computing the stationary moments of order 1. By Step 1 of Algorithm 2, $\boldsymbol{\Psi}^{(0,1)}$ satisfies the linear equation

$$
0 = \begin{bmatrix} -\overline{\alpha}_1 & \mathbb{E}[B_{12}] \\ \mathbb{E}[B_{21}] & -\overline{\alpha}_2 \end{bmatrix} \boldsymbol{\Psi}^{(0,1)} + \begin{bmatrix} \alpha_1\overline{\lambda}_1 \\ \alpha_2\overline{\lambda}_2 \end{bmatrix},
$$

yielding

$$
\boldsymbol{\Psi}^{(0,1)} = \frac{1}{\overline{\alpha}_1\overline{\alpha}_2 - \mathbb{E}[B_{12}]\mathbb{E}[B_{21}]} \begin{bmatrix} \alpha_1\overline{\lambda}_1\overline{\alpha}_2 + \alpha_2\overline{\lambda}_2\mathbb{E}[B_{12}] \\ \alpha_2\overline{\lambda}_2\overline{\alpha}_1 + \alpha_1\overline{\lambda}_1\mathbb{E}[B_{21}] \end{bmatrix}. \tag{91}
$$

Then, regarding Step 2, we find after some calculus

$$
\boldsymbol{\Psi}^{(1,0)} = \frac{1}{\overline{\alpha}_1\overline{\alpha}_2 - \mathbb{E}[B_{12}]\mathbb{E}[B_{21}]} \begin{bmatrix} \mu_1^{-1}\big(\alpha_1\overline{\lambda}_1\overline{\alpha}_2 + \alpha_2\overline{\lambda}_2\mathbb{E}[B_{12}]\big) \\ \mu_2^{-1}\big(\alpha_2\overline{\lambda}_2\overline{\alpha}_1 + \alpha_1\overline{\lambda}_1\mathbb{E}[B_{21}]\big). \end{bmatrix} \tag{92}
$$

We have thus found the stacked vector $\boldsymbol{\Psi}^{(1)}$. These expressions could also have been derived by sending $t \to \infty$ in the expressions of the transient moments $\boldsymbol{\Psi}_t^{(0,1)}$ and $\boldsymbol{\Psi}_t^{(1,0)}$, respectively.

To identify the second order stationary moments, we again go over the steps of Algorithm 2. Step 0 yields

$$\boldsymbol{\Psi}^{(0,2)} = \begin{bmatrix} 2\overline{\alpha}_1 & -2\mathbb{E}[B_{12}] & 0 \\ -\mathbb{E}[B_{21}] & \overline{\alpha}_1+\overline{\alpha}_2 & -\mathbb{E}[B_{12}] \\ 0 & -2\mathbb{E}[B_{21}] & 2\overline{\alpha}_2 \end{bmatrix}^{-1} \times \begin{bmatrix} \mathbb{E}[\lambda_1](2\alpha_1\overline{\lambda}_1+\mathbb{E}[B_{11}^2]) + \mathbb{E}[\lambda_2]\mathbb{E}[B_{12}^2] \\ \mathbb{E}[\lambda_1](\mathbb{E}[B_{11}]\mathbb{E}[B_{21}]+\alpha_2\overline{\lambda}_2) + \mathbb{E}[\lambda_2](\mathbb{E}[B_{22}]\mathbb{E}[B_{12}]+\alpha_1\overline{\lambda}_1) \\ \mathbb{E}[\lambda_1]\mathbb{E}[B_{21}^2] + \mathbb{E}[\lambda_2](2\alpha_2\overline{\lambda}_2+\mathbb{E}[B_{22}^2]) \end{bmatrix} \tag{93}$$

where the inverse may be explicitly computed in specific cases. For Step 1, we have, after some elementary matrix computations, that

$$\boldsymbol{\Psi}^{(1,1)} = \begin{bmatrix} \overline{\alpha}_1+\mu_1 & -\mathbb{E}[B_{12}] & 0 & 0 \\ -\mathbb{E}[B_{21}] & \overline{\alpha}_2+\mu_1 & 0 & 0 \\ 0 & 0 & \overline{\alpha}_1+\mu_2 & -\mathbb{E}[B_{12}] \\ 0 & 0 & -\mathbb{E}[B_{21}] & \overline{\alpha}_2+\mu_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbb{E}[\lambda_1^2]+\alpha_1\overline{\lambda}_1\mathbb{E}[Q_1]+\mathbb{E}[B_{11}]\mathbb{E}[\lambda_1] \\ \mathbb{E}[\lambda_1\lambda_2]+\alpha_2\overline{\lambda}_2\mathbb{E}[Q_1]+\mathbb{E}[B_{21}]\mathbb{E}[\lambda_1] \\ \mathbb{E}[\lambda_1\lambda_2]+\alpha_1\overline{\lambda}_1\mathbb{E}[Q_2]+\mathbb{E}[B_{12}]\mathbb{E}[\lambda_2] \\ \mathbb{E}[\lambda_2^2]+\alpha_2\overline{\lambda}_2\mathbb{E}[Q_2]+\mathbb{E}[B_{22}]\mathbb{E}[\lambda_2] \end{bmatrix}. \tag{94}$$

Finally, for Step 2 we have

$$\boldsymbol{\Psi}^{(2,0)} = \begin{bmatrix} 1/(2\mu_1) & 0 & 0 \\ 0 & 1/(\mu_1+\mu_2) & 0 \\ 0 & 0 & 1/(2\mu_2) \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \boldsymbol{\Psi}^{(1,1)}. \tag{95}$$

In line with earlier observations, the quasi-explicit results for the transient and stationary moments become involved for larger values of the order n.

D.4. **Higher order moments.** In this section, we provide some more detailed explicit matrices discussed Section 5, for the bivariate $d = 2$ setting for moments of order $n = 3$. For $n = 3$ we have $\mathfrak{D}(3,2) = 20$, and we consider the stacked vector (size 34)

$$\left(\boldsymbol{\Psi}_t^{(3)}, \boldsymbol{\Psi}_t^{(2)}, \boldsymbol{\Psi}_t^{(1)}\right)^\top,$$

which satisfies the ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \boldsymbol{\Psi}_t^{(3)} \\ \boldsymbol{\Psi}_t^{(2)} \\ \boldsymbol{\Psi}_t^{(1)} \end{bmatrix} = \boldsymbol{A}_3^{35\times 35} \begin{bmatrix} \boldsymbol{\Psi}_t^{(3)} \\ \boldsymbol{\Psi}_t^{(2)} \\ \boldsymbol{\Psi}_t^{(1)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{b}^{2\times 1} \\ \boldsymbol{0}^{32\times 1} \end{bmatrix}. \tag{96}$$

The matrix $\boldsymbol{F}_3^{34\times 34}$ is given by

$$\boldsymbol{F}_3^{34\times 34} = \begin{bmatrix} \boldsymbol{F}_2^{14\times 14} & \boldsymbol{0}^{14\times 20} \\ \boldsymbol{G}_3^{20\times 14} & \boldsymbol{H}_3^{14\times 14} \end{bmatrix}, \tag{97}$$

where

$$\boldsymbol{H}_3^{14\times 14} = \begin{bmatrix} \boldsymbol{M}^{(3,0)} & \mathbf{0}^{4\times 6} & \mathbf{0}^{4\times 6} & \mathbf{0}^{4\times 4} \\ \boldsymbol{K}^{(2,1)} & \boldsymbol{M}^{(2,1)} & \mathbf{0}^{6\times 6} & \mathbf{0}^{6\times 4} \\ \mathbf{0}^{6\times 4} & \boldsymbol{K}^{(1,2)} & \boldsymbol{M}^{(1,2)} & \mathbf{0}^{6\times 4} \\ \mathbf{0}^{4\times 4} & \mathbf{0}^{4\times 6} & \boldsymbol{K}^{(0,3)} & \boldsymbol{M}^{(0,3)} \end{bmatrix}, \quad \boldsymbol{G}_3^{20\times 14} = \begin{bmatrix} \boldsymbol{L}^{(0,3)} \\ \boldsymbol{L}^{(1,2)} \\ \boldsymbol{L}^{(2,1)} \\ \mathbf{0}^{4\times 20} \end{bmatrix}.$$

The elements of $\boldsymbol{G}_3^{20\times 20}$ require some more notation to describe, by introducing a number of sub-matrices. First, we have $\boldsymbol{L}^{(0,3)} = \begin{bmatrix}\boldsymbol{L}_{\lambda^1}^{(0,3)} & \mathbf{0}^{4\times 2} & \boldsymbol{L}_{\lambda^2}^{(0,3)} & \mathbf{0}^{4\times 7}\end{bmatrix}$, with

$$\boldsymbol{L}_{\lambda^1}^{(0,3)} = \begin{bmatrix} \mathbb{E}[B_{11}^3] & \mathbb{E}[B_{12}^3] \\ \mathbb{E}[B_{11}^2]\mathbb{E}[B_{21}] & \mathbb{E}[B_{12}^2]\mathbb{E}[B_{22}] \\ \mathbb{E}[B_{11}]\mathbb{E}[B_{21}^2] & \mathbb{E}[B_{12}]\mathbb{E}[B_{22}^2] \\ \mathbb{E}[B_{21}^3] & \mathbb{E}[B_{22}^3] \end{bmatrix},$$

$$\boldsymbol{L}_{\lambda^2}^{(0,3)} = \begin{bmatrix} 3\mathbb{E}[B_{11}^2] + 3\alpha_1\overline{\lambda}_1 & 3\mathbb{E}[B_{12}^2] & 0 \\ 2\mathbb{E}[B_{11}]\mathbb{E}[B_{21}] + \alpha_2\overline{\lambda}_2 & 2\mathbb{E}[B_{12}]\mathbb{E}[B_{22}] + \mathbb{E}[B_{11}^2] + 2\alpha_1\overline{\lambda}_1 & \mathbb{E}[B_{12}^2] \\ \mathbb{E}[B_{21}^2] & 2\mathbb{E}[B_{21}]\mathbb{E}[B_{11}] + \mathbb{E}[B_{22}^2] + 2\alpha_2\overline{\lambda}_2 & 2\mathbb{E}[B_{12}]\mathbb{E}[B_{22}] + \alpha_1\overline{\lambda}_1 \\ 0 & 3\mathbb{E}[B_{21}^2] & 3\mathbb{E}[B_{22}^2] + 3\alpha_2\overline{\lambda}_2 \end{bmatrix}.$$

Second, we have $\boldsymbol{L}^{(1,2)} = \begin{bmatrix}\boldsymbol{L}_{\lambda^1}^{(1,2)} & \mathbf{0}^{6\times 2} & \boldsymbol{L}_{\lambda^2}^{(1,2)} & \boldsymbol{L}_{Q^1\lambda^1}^{(1,2)} & \mathbf{0}^{6\times 3}\end{bmatrix}$, where

$$\boldsymbol{L}_{\lambda^1}^{(1,2)} = \begin{bmatrix} \mathbb{E}[B_{11}^2] & 0 \\ \mathbb{E}[B_{11}]\mathbb{E}[B_{21}] & 0 \\ \mathbb{E}[B_{21}^2] & 0 \\ 0 & \mathbb{E}[B_{12}^2] \\ 0 & \mathbb{E}[B_{12}]\mathbb{E}[B_{22}] \\ 0 & \mathbb{E}[B_{22}^2] \end{bmatrix}, \quad \boldsymbol{L}_{\lambda^2}^{(1,2)} = \begin{bmatrix} 2\mathbb{E}[B_{11}] & 0 & 0 \\ \mathbb{E}[B_{21}] & \mathbb{E}[B_{11}] & 0 \\ 0 & 2\mathbb{E}[B_{21} & 0 \\ 0 & 2\mathbb{E}[B_{12}] & 0 \\ 0 & \mathbb{E}[B_{22}] & \mathbb{E}[B_{12}] \\ 0 & 0 & 2\mathbb{E}[B_{22}] \end{bmatrix},$$

$$\boldsymbol{L}_{Q^1L^1}^{(1,2)} = \begin{bmatrix} \mathbb{E}[B_{11}^2] + 2\alpha_1\overline{\lambda}_1 & \mathbb{E}[B_{12}^2] & 0 & 0 \\ \mathbb{E}[B_{11}]\mathbb{E}[B_{21}] + \alpha_2\overline{\lambda}_2 & \mathbb{E}[B_{12}]\mathbb{E}[B_{22}] + \alpha_1\overline{\lambda}_1 & 0 & 0 \\ \mathbb{E}[B_{21}^2] & \mathbb{E}[B_{22}^2] + 2\alpha_2\overline{\lambda}_2 & 0 & 0 \\ 0 & 0 & \mathbb{E}[B_{11}^2] + 2\alpha_1\overline{\lambda}_1 & \mathbb{E}[B_{12}^2] \\ 0 & 0 & \mathbb{E}[B_{11}]\mathbb{E}[B_{21}] + \alpha_2\overline{\lambda}_2 & \mathbb{E}[B_{12}]\mathbb{E}[B_{22}] + \alpha_1\overline{\lambda}_1 \\ 0 & 0 & \mathbb{E}[B_{21}^2] & \mathbb{E}[B_{22}^2] + 2\alpha_2\overline{\lambda}_2 \end{bmatrix}.$$

Finally, $\boldsymbol{L}^{(2,1)} = \begin{bmatrix}\mathbf{0}^{6\times 7} & \boldsymbol{L}_{Q^1L^1}^{(2,1)} & \boldsymbol{L}_{Q^2}^{(2,1)}\end{bmatrix}$, where

$$\boldsymbol{L}_{Q^1L^1}^{(2,1)} = \begin{bmatrix} 2\mathbb{E}[B_{11}] & 0 & 0 & 0 \\ 2\mathbb{E}[B_{21}] & 0 & 0 & 0 \\ 0 & \mathbb{E}[B_{12}] & \mathbb{E}[B_{11}] & 0 \\ 0 & \mathbb{E}[B_{22}] & \mathbb{E}[B_{21}] & 0 \\ 0 & 0 & 0 & 2\mathbb{E}[B_{12}] \\ 0 & 0 & 0 & 2\mathbb{E}[B_{22}] \end{bmatrix}, \quad \boldsymbol{L}_{Q^2}^{(2,1)} = \begin{bmatrix} \alpha_1\overline{\lambda}_1 & 0 & 0 \\ \alpha_2\overline{\lambda}_2 & 0 & 0 \\ 0 & \alpha_1\overline{\lambda}_1 & 0 \\ 0 & \alpha_2\overline{\lambda}_2 & 0 \\ 0 & 0 & \alpha_1\overline{\lambda}_1 \\ 0 & 0 & \alpha_2\overline{\lambda}_2 \end{bmatrix}.$$

Appendix E. The Nearly Unstable Behavior: Proofs

Proof of Lemma 1. From the PDE in Eqn. (19) and the assumptions imposed,

$$\sum_{i=1}^{d}\big(\alpha s_i+\beta_i(\overline{s})-1\big)\frac{\mathrm{d}}{\mathrm{d}s_i}\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})=-\alpha\overline{\lambda}\sum_{i=1}^{d}s_i\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}),$$

upon substituting $z_1=\cdots=z_d=1$. Further observe that for any $i,j\in[d]$, we have

$$\frac{\mathrm{d}}{\mathrm{d}s_i}\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})=\mathbb{E}\big[-\lambda_i e^{-\boldsymbol{s}^\top\boldsymbol{\lambda}}\big]=\mathbb{E}\big[-\lambda_j e^{-\boldsymbol{s}^\top\boldsymbol{\lambda}}\big]=\frac{\mathrm{d}}{\mathrm{d}s_j}\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}),$$

since $\lambda_i\overset{d}{=}\lambda_j$, again by our assumptions on the parameters. Hence, we obtain the ODE

$$\frac{\mathrm{d}}{\mathrm{d}s_1}\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})=\frac{-\alpha\overline{\lambda}\overline{s}}{\alpha\overline{s}+\sum_{i=1}^{d}\beta_i(\overline{s})-d}\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})=:-f(s_1,\dots,s_d))\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}),\tag{98}$$

where we note that the choice of the index in the left-hand side is arbitrary. The solution to this ODE may be expressed as

$$\log\big(\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})\big)=-\int_0^{s_1}f(u,s_2,\dots,s_d)\mathrm{d}u+K,\quad K=\log(\mathcal{T}\{\boldsymbol{\lambda}\}(0,s_2,\dots,s_d)).\tag{99}$$

Since $\lambda_i\overset{d}{=}\lambda_j$, the Laplace transforms of the marginals satisfy, for any $s\in\mathbb{R}_+$,

$$\mathcal{T}\{\boldsymbol{\lambda}\}(s,0,\dots,0)=\mathcal{T}\{\boldsymbol{\lambda}\}(0,s,0,\dots,0)=\cdots=\mathcal{T}\{\boldsymbol{\lambda}\}(0,0,\dots,s).$$

We are then able to derive the joint Laplace transform of, say, $(\lambda_1,\lambda_2)^\top$ from Eqn. (99):

$$\begin{aligned}
&\mathcal{T}\{\boldsymbol{\lambda}\}(s_1,s_2,\dots,0)\\
&=\exp\Big(-\int_0^{s_1}f(u,s_2,\dots,0)\mathrm{d}u\Big)\mathcal{T}\{\boldsymbol{\lambda}\}(0,s_2,\dots,0)\\
&=\exp\Big(-\alpha\overline{\lambda}\int_0^{s_1}\frac{u+s_2}{\alpha(u+s_2)+\sum_{i=1}^{d}\beta_i(u+s_2)-d}\mathrm{d}u\Big)\\
&\quad\times\exp\Big(-\alpha\overline{\lambda}\int_0^{s_2}\frac{u}{\alpha u+\sum_{i=1}^{d}\beta_i(u)-d}\mathrm{d}u\Big)\\
&=\exp\Big(-\alpha\overline{\lambda}\big(\int_{s_2}^{s_1+s_2}\frac{v}{\alpha v+\sum_{i=1}^{d}\beta_i(v)-d}\mathrm{d}v+\int_0^{s_2}\frac{u}{\alpha u+\sum_{i=1}^{d}\beta_i(u)-d}\mathrm{d}u\big)\Big)\\
&=\exp\Big(-\alpha\overline{\lambda}\int_0^{s_1+s_2}\frac{v}{\alpha v+\sum_{i=1}^{d}\beta_i(v)-d}\mathrm{d}v\Big).
\end{aligned}\tag{100}$$

By symmetry, we can do this for any pair $(\lambda_i,\lambda_j)^\top$, with $i,j\in[d]$. Iterating the derivation in Eqn. (100), we obtain the full solution

$$\begin{aligned}
\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})&=\exp\Big(-\int_0^{s_1}f(u,s_2,\dots,s_d)\mathrm{d}u\Big)\mathcal{T}\{\boldsymbol{\lambda}\}(0,s_2,\dots,s_d)\\
&=\exp\Big(-\alpha\overline{\lambda}\int_0^{s_1+\cdots+s_d}\frac{u}{\alpha u+\sum_{i=1}^{d}\beta_i(u)-d}\mathrm{d}u\Big),
\end{aligned}$$

as desired. □

Proof of Theorem 3. The proof follows from the expression for $\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})$ in Eqn. (47), applying a Taylor expansion to the $\beta_i(\cdot)$, and computing the limit. Since the second moments of B_i exist, we have $\beta_i(u) = 1 - u\mathbb{E}[B_i] + \frac{1}{2}u^2\mathbb{E}[B_i^2] + o(u^2)$ as $u \downarrow 0$. Substituting $\boldsymbol{s}(1-\theta)$ as the argument in Eqn. (47) and the Taylor expansion of $\beta_i(\cdot)$, we obtain as $\theta \uparrow 1$, that

$$\begin{aligned}\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}(1-\theta)) &= \exp\Big(-\alpha\overline{\lambda}\int_0^{\overline{s}} \frac{u(1-\theta)}{\alpha u(1-\theta) + \sum_{i=1}^d \beta_i(u(1-\theta)) - d}(1-\theta)\mathrm{d}u\Big)\\ &= \exp\Big(-\overline{\lambda}\int_0^{\overline{s}} \frac{\alpha(1-\theta)}{\alpha - \sum_{i=1}^d \mathbb{E}[B_i] + \frac{u}{2}(1-\theta)\sum_{i=1}^d \mathbb{E}[B_i^2] + o(1-\theta)}\mathrm{d}u\Big)\\ &= \exp\Big(-\overline{\lambda}\int_0^{\overline{s}} \frac{1}{1 + \frac{u}{2\alpha}\sum_{i=1}^d \mathbb{E}[B_i^2] + o(1)}\mathrm{d}u\Big).\end{aligned}$$

Finally, by definition of σ and an elementary computation, we have

$$\lim_{\theta\uparrow 1}\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}(1-\theta)) = \lim_{\theta\uparrow 1}\exp\Big(-\overline{\lambda}\int_0^{\overline{s}} \frac{1}{1+u\sigma^{-1}+o(1)}\mathrm{d}u\Big) = \Big(\frac{\sigma}{\sigma+\overline{s}}\Big)^{\sigma\overline{\lambda}},$$

as claimed. □

Appendix F. Experiments with Transient Moments

This appendix demonstrates the numerical evaluation of the time-dependent moments featured in Section 3.1. From the joint transform characterization in Theorem 2, we compute the mixed moments, for any $t \geqslant 0$, $\tau > 0$ and any combination $i, j = 1, 2$,

$$\mathbb{E}[Q_i(t)\, Q_j(t+\tau)], \quad \mathbb{E}[\lambda_i(t)\, \lambda_j(t+\tau)], \tag{101}$$

as before using finite differences. It is noted that along the same lines objects of the type $\mathbb{E}[Q_i(t)\, \lambda_j(t+\tau)]$ can be evaluated, and in addition various types of auto-covariances and auto-correlations (cf. [7] for the auto-covariance in the market micro-structure setting). The joint transform characterization allows for efficient and fast computation of these cross-moments, also for large $t > 0$, which makes it practical in these settings.

To assess the efficiency and precision, we have conducted a numerical experiment with the same parameters as earlier in this subsection. To analyze the effect of the τ parameter in (101), in Figure 2 we fix $t = 1.5$ and plot the quantities of interest as functions of τ. The solid lines are the moments computed by applying FD to the joint transform, and the dotted lines represent the results from the MC method (based on 10^4 runs). We see that MC performs increasingly poorly as τ increases, in particular for the population processes $Q_i(\cdot)$, which is due to the fact that there are more events (i.e., arrivals and departures) for larger τ, and hence more variation. Furthermore, the different shapes in the plots indicate that the effect of τ on the specific cross-moment depends on the chosen parameters.

We conclude our numerical section by an experiment that quantifies the effect of the initial values. In Theorem 1, we characterized the joint transform with the processes being initialized at $\boldsymbol{Q}(t_0) = (Q_1(t_0), Q_2(t_0)) = (q_{1,0}, q_{2,0}) \in \mathbb{N}^2$ and $\boldsymbol{\lambda}(t_0) = (\lambda_1(t_0), \lambda_2(t_0)) = (\lambda_{1,0}, \lambda_{2,0}) \in \mathbb{R}_+^2$ for some $t_0 > 0$. By applying FD, we can compute the moments of our interest for any initial

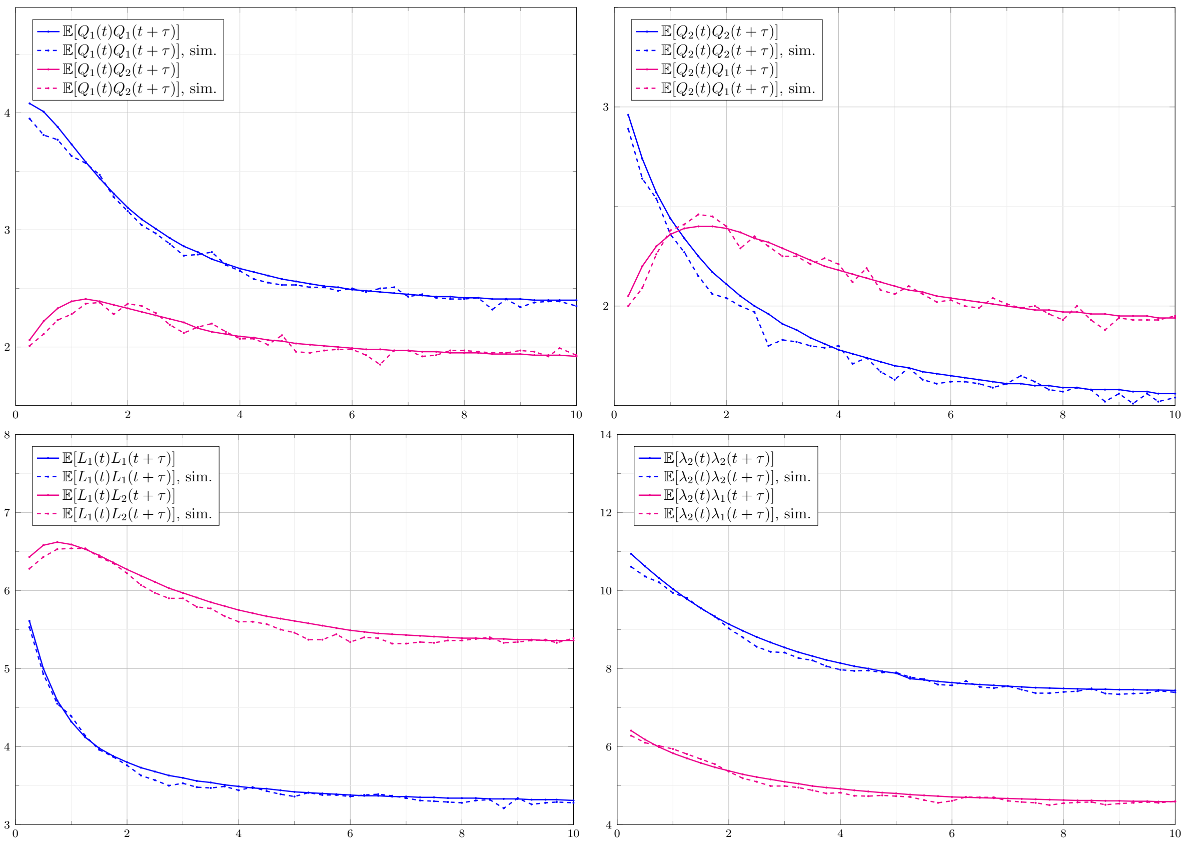


Figure 2. Computation of cross-moments for $t = 1.5$ and $\tau \in [0, 10]$ using the joint transform characterization (solid lines) compared to Monte Carlo simulated averages (dashed lines).

value, for instance

$$\mathbb{E}[Q_i(t) \,|\, Q_i(t_0) = q_{i,0}], \quad \mathbb{E}[\lambda_i(t) \,|\, \lambda_i(t_0) = \lambda_{i,0}], \tag{102}$$

with $i = 1, 2$, where $q_{i,0} \in \mathbb{N}$ and $\lambda_{i,0} \in \mathbb{R}_+$. In our experiment we focus on the moments of $Q_1(\cdot)$ and $\lambda_1(\cdot)$, studying the effect of three different choices of $q_{i,0}$ and $\lambda_{i,0}$ on the first moments and variances as a function of t. Note that a different value of $Q_i(t_0) = q_{i,0}$ will not influence $\lambda_j(\cdot)$, since the population processes do not directly affect the intensity processes, but due to mutual excitation, the values $\lambda_1(t_0) = \lambda_{1,0}$ and $\lambda_2(t_0) = \lambda_{2,0}$ do matter. When computing $\mathbb{E}[Q_1(t)|Q_1(t_0) = q_{1,0}]$, we leave $\lambda_i(t_0) = \lambda_i(0) = \overline{\lambda}_i$ for $i = 1, 2$, and we only change $q_{1,0}$. Similarly, when computing $\mathbb{E}[\lambda_1(t)|\lambda_1(t_0) = \lambda_{1,0}]$, we leave $\lambda_2(t_0) = \lambda_2(0) = \overline{\lambda}_2$ and only change $\lambda_{1,0}$.

Figure 3 shows the expectations and variances, where we introduced the compact notation $\mathbb{E}[Q_1(t) \,|\, q_{1,0}] = \mathbb{E}[Q_1(t) \,|\, Q_1(t_0) = q_{1,0}]$ and $\mathbb{E}[\lambda_1(t)|\lambda_{1,0}] = \mathbb{E}[\lambda_1(t)|\lambda_1(t_0) = \lambda_{1,0}]$, and similar notation for the variances. For the moments of $Q_1(t)$, we observe a vertical shift of the plots, which is expected since the arrived individuals depart independently and according to the same distribution. For the moments of $\lambda_1(t)$, we see that the effect of the $\lambda_{1,0}$-value is substantial. For both the mean and the variance there is convergence to their respective steady-state values.

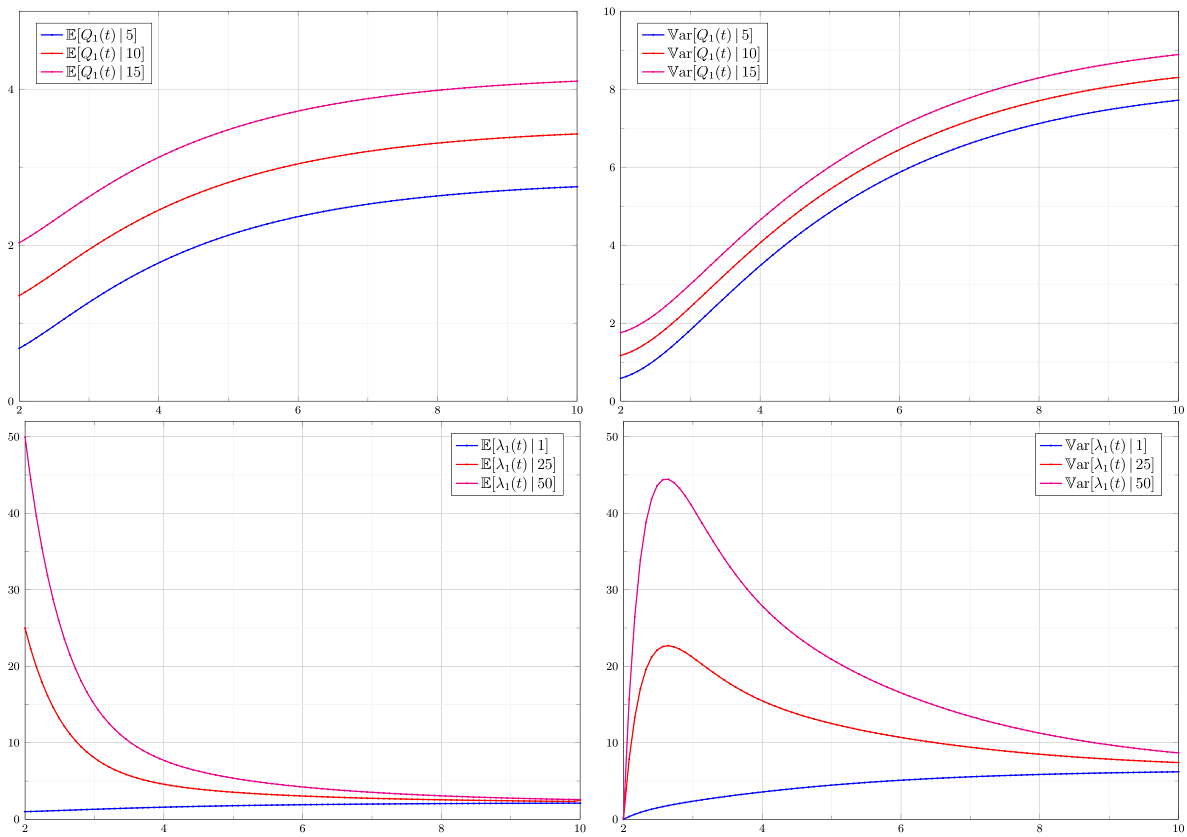


Figure 3. Computation of expected values and variances of $Q_1(t)$ and $\lambda_1(t)$, with different initial values at $t_0 = 2$, with $t \in [t_0, 10]$, using the joint transform characterization.