Mixture Modeling for Temporal Point Processes with Memory

Xiaotian Zheng, Athanasios Kottas and Bruno Sansó*

May 16, 2023

Abstract

We propose a constructive approach to building temporal point processes that incorporate dependence on their history. The dependence is modeled through the conditional density of the duration, i.e., the interval between successive event times, using a mixture of first-order conditional densities for each one of a specific number of lagged durations. Such a formulation for the conditional duration density accommodates high-order dynamics, and it thus enables flexible modeling for point processes with memory. The implied conditional intensity function admits a representation as a local mixture of first-order hazard functions. By specifying appropriate families of distributions for the first-order conditional densities, with different shapes for the associated hazard functions, we can obtain either self-exciting or self-regulating point processes. From the perspective of duration processes, we develop a method to specify a stationary marginal density. The resulting model, interpreted as a dependent renewal process, introduces high-order Markov dependence among identically distributed durations. Furthermore, we provide extensions to cluster point processes. These can describe duration clustering behaviors attributed to different factors, thus expanding the scope of the modeling framework to a wider range of applications. Regarding implementation, we develop a Bayesian approach to inference and model checking. We investigate point process model properties analytically, and illustrate the methodology with both synthetic and real data examples.

Keywords: Bayesian hierarchical models; Cluster point processes; Copulas; Dependent point processes; Mixture transition distribution models; Self-exciting processes.

*Xiaotian Zheng (xzheng@uow.edu.au) is Research Fellow in the School of Mathematics and Applied Statistics, University of Wollongong, Australia, and Athanasios Kottas (thanos@soe.ucsc.edu) and Bruno Sansó (bruno@soe.ucsc.edu) are Professors in the Department of Statistics, University of California, Santa Cruz. This research is part of the Ph.D. dissertation of X. Zheng, completed at University of California, Santa Cruz.

1 Introduction

Temporal point processes are stochastic models for sequences of random events that occur in continuous time, with irregular durations, i.e., intervals between successive arrival times. Throughout this article, event time and arrival time will be used interchangeably for the occurrence time of an event. Data corresponding to point patterns are common in a wide range of applications, such as earthquake occurrences (Ogata, 1988), recurrent events (Cook et al., 2007), financial high frequency trading and orders (Hautsch, 2011), and neural spike trains (Tang and Li, 2021), to name a few. For many point patterns, it is believed that occurrence of a future event depends on the past. This motivates the use of point processes with memory, for example, the Hawkes process (Hawkes, 1971) with full memory, or renewal processes with lagged dependence. The goal of this article is to propose a modeling framework for point processes with high-order memory, relaxing the independent duration assumption in the traditional renewal process, and including the ability to model duration clustering behaviors that are present in applications such as health care (Yang et al., 2018), climatology (Cowpertwait, 2001), and finance (O'Hara, 1995).

A popular way to model point process dependence is by specifying the process conditional intensity, namely, the instantaneous event rate conditional on the process history. Under this approach, the Hawkes process has been used extensively in the literature. The conditional intensity of the Hawkes process is decomposed into a baseline intensity and a triggering component. The triggering component is commonly specified through an excitation function, such as an exponential or power law kernel. As a result, a new event causes a jump in the conditional intensity, and thus the Hawkes process is said to be a self-exciting point process. We refer to Reinhart (2018) and references therein for a review.

This article explores an alternative approach for modeling point processes with memory. Specifically, we consider models for the durations. As these correspond to discrete-time stochastic processes, a time series model for dependent durations induces conditional densities on the arrival times. We refer to the conditional densities as conditional arrival densities, and notice that they uniquely determine the distribution of the resulting point process (Daley and Vere-Jones, 2003). In fact, a point process can be equivalently characterized by its conditional intensity or the conditional arrival densities. The latter approach benefits from the vast literature on conditional density modeling. In contrast, building models for conditional intensities requires mathematical validation of the proposed intensity function to obtain a well-defined point process model. Regarding inference, both approaches involve integration to obtain the corresponding normalizing term of the likelihood function. The conditional intensity approach requires integration of the intensity function, while the integration required for duration models is more efficient as it involves a conditional cumulative distribution function (c.d.f.). Thus, constructing point processes using duration models, usually coupled with a limited memory assumption, can be computationally attractive. In Section 2, we provide more detailed discussion of the two approaches.

Statistical models for durations date back at least to Wold (1948) who proposed a first-order Markov chain with an additive model representation. Subsequent developments (Jacobs and Lewis, 1977; Gaver and Lewis, 1980) investigate specific families for the duration process stationary marginal distribution. Since durations are positive-valued, a structure with an additive error process is in general restrictive. A popular class of models in finance is built from the autoregressive conditional duration (ACD) structure (Engle and Russell, 1998). The ACD model assumes independent and identically distributed (i.i.d.) multiplicative errors for the durations, with each multiplicative factor modeled as a linear function of the past factors and durations. Extensions of this class of models provide additional flexibility through the multiplicative factor specification or the error distribution choice. We refer to Pacurar (2008) and Bhogal and Thekke Variyam (2019) for comprehensive reviews. For these models, the conditional intensity function is obtained by scaling the baseline hazard function with multiplicative factors. The baseline hazard corresponds to the error distribution, typically chosen within a parametric family. A restriction of ACD models is their limited capacity to handle non-linear dynamics. Regarding computation, the ACD model structure complicates inference when the assumption of high-order memory is necessary. In particular, estimating the correlated multiplicative factors can be difficult and may require approximations (Strickland et al., 2006).

A different approach to modeling duration dependence involves mixture transition distribution (MTD) models (Le et al., 1996), which describe the transition density of a time series as a weighted combination of first-order conditional densities for each one of a specified number of lags. Hassan and Lii (2006) propose a bivariate MTD model for the joint conditional distribution of the duration and a continuous mark, i.e., a random variable associated with the point events. Hassan and El-Bassiouni (2013) extend the model to include a discrete mark. However, this work does not investigate point process properties, such as stationarity, and it requires certain families of distributions for the duration and mark, which can be practically restrictive. Hassan and Lii (2006) point out that the choice of parameterization is non-trivial to ensure model stability and prediction capability.

In this article, we introduce a class of temporal point process models that builds on the idea of modeling duration process dynamics with MTD models. In general, it is difficult to model dependent, positive-valued durations using traditional high-order autoregressive models without transforming the durations. The challenges include model inference under a constrained, possibly high-dimensional parameter space. For example, coefficients need to be restricted to avoid negative-valued durations, and practical implementation for stationarity conditions is difficult. The aforementioned work that uses MTD models attempts to handle the former issue, albeit under restrictive structures. A key contribution of the present article is the development of an MTD point process (MTDPP) constructive framework that provides flexible modeling of high-order dynamics for the duration process, without parameter constraints. The framework allows modeling for various types of practically relevant point patterns, such as those with self-excitation or self-regulation effects. In addition, it provides efficient implementation of model inference. The MTDPP likelihood evaluation grows linearly with the number of events, thus delivering computational scalability, especially for large point patterns with high-order memory.

Within the MTDPP framework, we provide easily-implemented conditions to construct point processes given a pre-specified family for the duration marginal distribution, and obtain a limit result for the mean value function analogous to that for renewal processes. The resulting class of models has identically distributed, high-order dependent durations, and can be interpreted as a class of dependent renewal processes. This relaxes the assumption of independent durations that may be restrictive in practice.

Moreover, we develop a model extension for duration clustering, using a two-component mixture model for the conditional duration density. In particular, one component of the mixture corresponds to an independent duration model that accounts for external factors. The other component is an MTDPP that models self-excitation. Point patterns of this type can be found, for instance, in hospital emergency department visits of patients, where long durations may be observed between clusters of multiple visits in short bursts (Yang et al., 2018), and in financial markets where fluctuation can be caused by either external or internal processes (Filimonov and Sornette, 2012). The model extension accounts for the possibility of two different factors that may drive the point process dynamics.

The rest of the article is organized as follows. Section 2 introduces the MTDPP framework, including study of model properties, approaches to constructing various types of MTDPP models, and the extension to cluster point processes. Section 3 develops the Bayesian model formulation, Markov chain Monte Carlo (MCMC) inference, and a model validation method. In Section 4, we illustrate the proposed methodology with synthetic and real data examples. Finally, Section 5 concludes with a summary and discussion.

2 Temporal MTD point processes

We consider a temporal point process $N(t)$ defined on the positive half-line $\mathbb{R}^+$, where $N(t) = \sum_{i\geq 1} \mathbb{1}_{\{t_i \leq t\}}$ is a right-continuous, integer-valued function, and $t_1, t_2, \dots \in \mathbb{R}^+$ denote the event times. A temporal point process is usually modeled by its conditional intensity, $\lambda^*(t) \equiv \lambda(t \,|\, \mathcal{H}_t) = \lim_{dt\to 0} E[dN(t) \,|\, \mathcal{H}_t]/dt$, where $dN(t) = N(t+dt) - N(t)$, and $\mathcal{H}_t$ is the process history up to but not including t. The point process has memory if $\lambda^*(t)$ depends on the process history. A Poisson process is an example of a memoryless process. A renewal process has limited memory, in particular, $\mathcal{H}_t = t_{N(t)}$, where $t_{N(t)}$ is the most recent arrival time before t. In contrast, the evolution of a Hawkes process depends on the entire past. Given an observed point pattern $\{t_i\}_{i=1}^n$ over $(0,T)$, the likelihood is

$$p(t_1, \dots, t_n) = \left(\prod_{i=1}^n \lambda^*(t_i) \right) \exp\left(- \int_0^T \lambda^*(t) dt \right), \tag{1}$$

where the last component of (1) corresponds to the normalizing term, which is typically analytically intractable, especially when $\lambda^*(t)$ has a complicated form.

An alternative way to characterize the point process probability structure is to use the collection of conditional arrival densities, denoted as $p_i^*(t) \equiv p_i(t \,|\, \mathcal{H}_t)$, supported on (t_{i-1}, ∞), with associated conditional survival functions $S_i^*(t) = 1 - \int_{t_{i-1}}^t p_i^*(u) du$. When $i = 1$, $p_1^*(t) \equiv p_1(t)$ and $S_1^*(t) = 1 - \int_0^t p_1^*(u) du$, where p_1 is the marginal density of the first event time. Now, the likelihood for point pattern $\{t_i\}_{i=1}^n$ observed in $(0,T)$ is given by

$$p(t_1, \dots, t_n) = \left(\prod_{i=1}^n p_i^*(t_i) \right) \left(1 - \int_{t_n}^T p_{n+1}^*(u) du \right). \tag{2}$$

Similar to (1), the last component of (2) defines the likelihood normalizing term, i.e., the probability of no events occurring in the interval (t_n, T). Since the normalizing term corresponds to a conditional c.d.f., it may be available in closed-form for particular model formulations for the conditional arrival densities.

Using the collection of conditional densities p_i^* and survival functions S_i^*, we can define the hazard functions as $\lambda_i^*(t) = p_i^*(t)/S_i^*(t)$, for $i = 1, \ldots, n$. The hazard function is naturally interpreted as the conditional instantaneous event rate. Consequently, given the set of arrival times, we can write the conditional intensity of the process as $\lambda^*(t) = \lambda_i^*(t),\ t_{i-1} < t \leq t_i,\ 1 \leq i \leq n$. Since $p_i^*(t) = \lambda_i^*(t) \exp(-\int_{t_{i-1}}^{t} \lambda_i^*(u) du)$, we can use the form in (2) to recover the likelihood in (1).

Although there is an one-to-one correspondence between modeling the conditional intensity and the conditional arrival densities, the computational cost may be different. For instance, for the Hawkes process, the conditional intensity involves the sum of the excitation function over all points from the past, which poses challenges to model estimation (Veen and Schoenberg, 2008), and the likelihood evaluation cost grows quadratically with the number of observed points. Point process models defined using conditional arrival densities typically assume limited memory, with an autoregressive structure on the durations. The resulting likelihood based on (2) is similar to that of an autoregressive time series, with an extra term that corresponds to a survival function. In general, the likelihood formulation in (2) facilitates model-based inference for temporal point processes with memory.

2.1 Conditional duration density

Consider an ordered sequence of arrival times $0 = t_0 < t_1 < \cdots < t_n < T$, and denote the durations by $x_i = t_i - t_{i-1}$, for $i = 1, \ldots, n$. The memory of the process is modeled by specifying an MTD structure for the conditional duration densities. In particular, the density of x_i conditional on the past L durations is modeled as a weighted combination of first-order transition densities, each of which depends on a specific past duration, i.e., $f(x_i \mid x_{i-1}, \ldots, x_1) = \sum_{l=1}^{L} w_l\, f_l(x_i \mid x_{i-l})$, where $w_l \geq 0$, for all l, and $\sum_{l=1}^{L} w_l = 1$. Transforming the conditional density of x_i to that for $t_i = t_{i-1} + x_i$, for every i, creates conditional arrival densities that uniquely determine the point process. The construction is motivated above for durations x_i with $i > L$. The formal MTDPP definition is given as follows.

Definition 1. Let $N(t)$ be a temporal point process defined on $\mathbb{R}^+$ with event arrival times $t_1, t_2, \cdots \in \mathbb{R}^+$. Denote by $f^*(t - t_{N(t)}) \equiv f(t - t_{N(t)} \,|\, \mathcal{H}_t)$ the conditional duration density. Then $N(t)$ is said to be an *MTD point process* if (i) $t \sim f_0$ for $N(t) = 0$; (ii) for $1 \leq N(t) \leq L - 1$, the conditional duration density

$$f^*(t - t_{N(t)}) = \sum_{l=1}^{N(t)-1} w_l \, f_l(t - t_{N(t)} \,|\, t_{N(t)-l+1} - t_{N(t)-l}) + (1 - \sum_{r=1}^{N(t)-1} w_r) f_{N(t)}(t - t_{N(t)} \,|\, t_1); \tag{3}$$

(iii) for $N(t) \geq L$, the conditional duration density

$$f^*(t - t_{N(t)}) = \sum_{l=1}^{L} w_l \, f_l(t - t_{N(t)} \,|\, t_{N(t)-l+1} - t_{N(t)-l}). \tag{4}$$

In both (3) and (4), the weights $w_l \geq 0$, for $l = 1, \ldots, L$, with $\sum_{l=1}^{L} w_l = 1$.

Remark 1. The marginal density f_0 and the conditional density $f^*(t - t_{N(t)})$ define the conditional arrival densities p_i^* for point pattern $\{t_i\}_{i=1}^n$, by taking $p_1^*(t) = f_0(t)$ and $p_i^*(t) = f^*(t - t_{i-1})$, $t > t_{i-1}$, for $i = 2, \ldots, n$. Thus, specification of densities $f_0(t)$ and $f^*(t - t_{N(t)})$ suffices to characterize the probability structure of the resulting MTDPP.

Remark 2. The two different expressions (3) and (4) for the conditional duration density allow us to study stationarity conditions for the MTDPP (Section 2.2). Regarding inference, Equation (4) is the relevant expression, since we work with a conditional likelihood. For brevity, we will use (4) to discuss model properties throughout the rest of the article.

The specification of the conditional density $f^*(t - t_{N(t)})$ involves the first-order conditional density f_l, for $l = 1, \ldots, L$. Following Zheng et al. (2022), we build f_l from a bivariate positive-valued random vector (U_l, V_l) with joint density f_{U_l,V_l} and marginals f_{U_l}

and f_{V_l}, by taking $f_l = f_{U_l|V_l}$ as the conditional density of U_l given V_l. In general, there are two strategies to define the joint density f_{U_l,V_l}, one through specific marginal densities, and the other through a pair of compatible conditional densities (Arnold et al., 1999). The two conditional densities $f_{U_l|V_l}$ and $f_{V_l|U_l}$ are said to be compatible if there exists a bivariate density with its conditionals given by $f_{U_l|V_l}$ and $f_{V_l|U_l}$. We note that each strategy has its own benefits depending on the modeling objective. In Section 2.3, we illustrate construction of the conditional densities f_l with various examples for different goals.

An important consequence of using the MTD model for the conditional duration density is a mixture formulation for the implied conditional intensity $\lambda^*(t) \equiv h^*(t - t_{N(t)}) = f^*(t - t_{N(t)})/S^*(t - t_{N(t)})$, where $h^*(t - t_{N(t)})$ and $S^*(t - t_{N(t)})$ are the hazard and survival function, respectively, associated with $f^*(t - t_{N(t)})$. Let h_l and S_l be the hazard and survival function associated with f_l. Then, we can write the conditional intensity $\lambda^*(t)$ as

$$\lambda^*(t) = \sum_{l=1}^{L} w_l^*(t)\, h_l(t - t_{N(t)} \,|\, t_{N(t)-l+1} - t_{N(t)-l}), \tag{5}$$

with weights $w_l^*(t) = w_l\, S_l(t - t_{N(t)} \,|\, t_{N(t)-l+1} - t_{N(t)-l})/S^*(t - t_{N(t)})$, where $S^*(t - t_{N(t)}) = \sum_{l=1}^{L} w_l\, S_l(t - t_{N(t)} \,|\, t_{N(t)-l+1} - t_{N(t)-l})$. Note that $w_l^*(t) \geq 0$ and $\sum_{l=1}^{L} w_l^*(t) = 1$ for all t. The time-dependent weights, $w_l^*(t)$, provide local adjustment, and thus the flexibility to accommodate a wide range of conditional intensity shapes.

In addition to model flexibility, the mixture formulation of $\lambda^*(t)$ guides modeling choice. Each mixture component h_l is a first-order hazard function. If we select f_l such that $h_l \leq B_l$, for constant $B_l > 0$, and for all l, then $\lambda^*(t) \leq \sum_{l=1}^{L} w_l^*(t)\, B_l$, for every t. Similarly, we can find a lower bound for $\lambda^*(t)$. For both cases, if $h_l \to B$ as $t \to \infty$ for all l, we have that $\lambda^*(t) \to B$ as $t \to \infty$. On the other hand, if one of the component hazard functions $h_l \to \infty$ as $t \to \infty$, then $\lambda^*(t) \to \infty$. Moreover, choosing f_l such that h_l has certain shapes results in particular types of point processes. A point process is said to be self-exciting if a new arrival causes the conditional intensity to jump, and is called self-regulating (or self-correcting) if a new arrival causes the conditional intensity to drop. If h_l monotonically decreases, for all l, the resulting MTDPP is self-exciting; see Section 2.3 for details.

2.2 Model properties

We first investigate stationarity for MTDPPs. We focus on conditions for first-order strict stationarity, such that the MTDPP has a stationary marginal density, f_X, for the duration process. The constructive approach to build f_l as the conditional density $U_l \mid V_l$ based on random vector (U_l, V_l) allows us to obtain a stationary marginal density f_X, using the approach in Zheng et al. (2022). We summarize the conditions in the following proposition.

Proposition 1. *Consider an MTD point process $N(t)$ with event arrival times $t_1, t_2, \dots \in \mathbb{R}^+$. Let $\{X_i : i \geq 1\}$ be the duration process, with $x_1 = t_1$, and $x_i = t_i - t_{i-1}$, for $i \geq 2$. The duration process has a stationary marginal density f_X if: (i) $t \sim f_X$ for $N(t) = 0$; (ii) the density f_l in (3) and (4) is taken to be the conditional density $f_{U_l|V_l}$ of a bivariate positive-valued random vector (U_l, V_l) with marginal densities f_{U_l} and f_{V_l}, such that $f_{U_l}(x) = f_{V_l}(x) = f_X(x)$, for all $x \in \mathbb{R}^+$ and for all l.*

We refer to the class of MTDPPs that satisfies the conditions in Proposition 1 as stationary MTDPPs. Compared to renewal processes that have i.i.d. durations, stationary MTDPPs can be interpreted as dependent renewal processes, where the durations are identically distributed, and Markov-dependent, up to L-order. In fact, the independence assumption in classical renewal processes is often unrealistic (Coen et al., 2019). For example, in reliability engineering, times to failure between component replacements can be correlated (Modarres et al., 2017). Another example involves the analysis of the recurrence interval distribution for extreme events, which is illustrated in Section 4.2.

For the class of stationary MTDPPs, it is possible to obtain a limit result analogous to that of renewal processes. In renewal theory, the rate of renewals (e.g., component replacement) in the long run corresponds to the rate at which $N(t)$ goes to infinity, i.e., $\lim_{t\to\infty} N(t)/t$. The following theorem summarizes the limit result for stationary MTDPPs.

Theorem 1. *Consider an MTD point process $N(t)$ such that its duration process has stationary marginal density f_X with finite mean $\mu > 0$ and finite variance. It holds that, as $t \to \infty$, $N(t)/t \to 1/\mu$ almost surely.*

Similar to the classical renewal theorem, Theorem 1 shows the average renewal rate, the difference being that the MTDPP allows dependence among waiting times between renewals. Another rate of general interest concerns the mean-value function $m(t) = E[N(t)]$, that is, $\lim_{t\to\infty} m(t)/t$. The MTDPP function $m(t)$ involves integration with respect to the probability distribution of the point process, and thus, in general, it is not analytically available. However, a useful upper bound for the rate $\lim_{t\to\infty} m(t)/t$ can be obtained for MTDPPs with bounded component hazard functions, as summarized below.

Proposition 2. *Consider an MTD point process $N(t)$ with conditional intensity given by (5), such that, for all l, the component hazard functions satisfy $h_l \leq B_l$. Then, $\lim_{t\to\infty} m(t)/t \leq \sum_{l=1}^{L} w_l B_l$.*

In the context of using stationary MTDPPs for modeling renewals, Proposition 2 implies that the expected average renewal rate is no larger than a convex sum of the hazard rate upper bounds. In summary, the results of this section can guide modeling choices, combined with the approaches to constructing MTDPPs presented in Section 2.3.

2.3 Construction of MTD point processes

We provide guidance to construct MTDPPs, focusing on the conditional density f_l. As discussed in Section 2.1, we derive f_l from a bivariate density f_{U_l,V_l}, which can be specified through compatible conditionals $f_{U_l|V_l}$ and $f_{V_l|U_l}$, or through marginals f_{U_l} and f_{V_l}. The former is particularly useful when the objective is to construct self-exciting or self-regulating MTDPPs, by choosing $f_{U_l|V_l}$ such that its associated hazard function is monotonically decreasing or increasing, respectively. We illustrate this approach in Example 1.

In light of Proposition 1, the strategy of constructing MTDPPs through pre-specified marginals is natural for modeling dependent renewal processes. This strategy is also useful when interest lies in the shape of the marginal hazard function. For example, Grammig and Maurer (2000) point out that it may be more appropriate to consider non-monotonic hazard functions for modeling financial duration processes. We implement this MTDPP construction approach using bivariate copula functions for f_{U_l,V_l}, illustrated in Example 2.

Example 1: Self-exciting and self-regulating MTDPPs

We derive MTDPPs based on bivariate Lomax distributions. The Lomax distribution is a shifted version of the Pareto Type I distribution, denoted as $P(u \,|\, b, a) = ab^{-1}(1 + ub^{-1})^{-(a+1)}$, where $a > 0$ is the shape parameter, and $b > 0$ is the scale parameter. In this example, we derive a new pair of compatible conditionals, based on the pair of Lomax conditionals in Arnold et al. (1999). The definition is given in the following proposition.

Proposition 3. *Consider a bivariate Lomax random vector* (X, Y) *with density* $f_{X,Y}(x, y) \propto (\lambda_0 + \lambda_1 x + \lambda_2 y)^{-(\alpha+1)}$. *Let* $(U, V) = (\alpha X, \alpha Y)$. *Then, the bivariate random vector* (U, V) *has conditionals* $f_{U|V}(u|v) = P(u \,|\, \lambda_1^{-1}(\alpha\lambda_0 + \lambda_2 v), \alpha)$ *and* $f_{V|U}(v|u) = P(v \,|\, \lambda_2^{-1}(\alpha\lambda_0 + \lambda_1 u), \alpha)$*, and marginals* $f_U(u) = P(u \,|\, \lambda_1^{-1}\alpha\lambda_0, \alpha - 1)$ *and* $f_V(v) = P(v \,|\, \lambda_2^{-1}\alpha\lambda_0, \alpha - 1)$*, where* $\lambda_0 > 0$*,* $\lambda_1 > 0$*,* $\lambda_2 > 0$*, and* $\alpha > 1$.

Since (X, Y) is scaled by α, we refer to the distribution of (U, V) as the bivariate scaled-Lomax distribution. The difference with the original Lomax distribution is that the shape parameter of the scaled-Lomax distribution is part of the scale parameter.

To construct an MTDPP, we start with bivariate scaled-Lomax densities f_{U_l,V_l} with parameters $\alpha_l, \lambda_{0l}, \lambda_{1l}, \lambda_{2l}$. We simplify the parameterization by setting $\lambda_l = \lambda_{1l} = \lambda_{2l}$, and letting $\phi_l = \lambda_{0l}/\lambda_l$, which yields $f_{U_l|V_l}(u|v) = P(u \,|\, \alpha_l \phi_l + v, \alpha_l)$, where $\phi_l > 0$ and $\alpha_l > 1$, for all l. Taking $f_l \equiv f_{U_l|V_l}$, we obtain the conditional duration density,

$$f^*(t - t_{N(t)}) \;=\; \sum_{l=1}^{L} w_l \, P(t - t_{N(t)} \,|\, \alpha_l \phi_l + t_{N(t)-l+1} - t_{N(t)-l}, \alpha_l). \tag{6}$$

We complete the scaled-Lomax MTDPP construction with $f_0(t) = P(t \,|\, \alpha_1 \phi_1, \alpha_1 - 1)$.

Based on Proposition 1, if $\alpha_l = \alpha$ and $\phi_l = \phi$, for all l, the model has stationary duration density $P(\alpha\phi, \alpha - 1)$. The next result describes the limiting behavior of the stationary scaled-Lomax MTDPP conditional duration distribution.

Proposition 4. *Consider the stationary scaled-Lomax MTDPP with marginal duration density $P(\alpha\phi, \alpha - 1)$. As $\alpha \to \infty$, the conditional duration distribution converges in distribution to the exponential distribution with rate parameter ϕ^{-1}.*

According to (5), the conditional intensity of the scaled-Lomax MTDPP can be expressed as $\lambda^*(t) = \sum_{l=1}^{L} w_l^*(t) \, \{\phi_l + \alpha_l^{-1}(t - t_{N(t)} + t_{N(t)-l+1} - t_{N(t)-l})\}^{-1}$. For each l, the lth component of the conditional intensity is bounded above by ϕ_l^{-1}. Thus, $\lambda^*(t) \leq \sum_{l=1}^{L} w_l^*(t) \, \phi_l^{-1}$, for any t, and, using Proposition 2, $\lim_{t \to \infty} m(t)/t \leq \sum_{l=1}^{L} w_l \, \phi_l^{-1}$.

Finally, we note that if we remove α from the scale parameter component in (6), such that $f_l = P(t - t_{N(t)} \,|\, \phi_l + t_{N(t)-l+1} - t_{N(t)-l}, \alpha_l)$ and $f_0(t) = P(t \,|\, \phi_1, \alpha_1 - 1)$, then f_l corresponds to the bivariate Lomax distribution of Arnold et al. (1999). The resulting point process is referred to as the Lomax MTDPP. Since the hazard function of a Lomax distribution is monotonically decreasing, both the scaled-Lomax MTDPP and Lomax MTDPP are self-exciting point processes. A self-regulating MTDPP can be constructed through compatible conditionals associated with monotonically increasing hazard functions, such as gamma conditionals; see Arnold et al. (1999) for relevant bivariate distributions.

Example 2: Dependent renewal MTDPPs

Motivated by Proposition 1, we can select a stationary density f_X, and take $f_{U_l} = f_{V_l} = f_X$, for all l. Given the desired marginals, what remains is to specify the joint density f_{U_l,V_l} to obtain $f_{U_l|V_l}$. In this example, we introduce the idea of specifying a bivariate copula function $C : [0,1]^2 \to [0,1]$ to build f_{U_l,V_l}, which provides a general scheme to construct MTDPPs given a stationary marginal f_X.

Let F_{U_l,V_l} be the joint c.d.f. of the random vector (U_l, V_l), and denote by F_{U_l}, F_{V_l} the corresponding marginal c.d.f.s. Given F_{U_l} and F_{V_l}, there exists a unique copula C_l such that $F_{U_l,V_l}(u,v) = C_l(F_{U_l}(u), F_{V_l}(v))$, and the joint density f_{U_l,V_l} is given by $c_l(u,v)f_{U_l}(u)f_{V_l}(v)$, where $c_l(u,v) = \partial^2 C(F_{U_l}(u), F_{V_l}(v))/(\partial F_{U_l}\partial F_{V_l})$ is the copula density (Sklar, 1959). Hence, based on a marginal duration density f_X and a copula C_l, we have $f_l(u) \equiv f_{U_l|V_l}(u \,|\, v) = c_l(u,v)f_X(u)$. The conditional duration density of the resulting MTDPP is

$$f^*(t - t_{N(t)}) \;=\; \sum_{l=1}^{L} w_l \, c_l(t - t_{N(t)}, t_{N(t)-l+1} - t_{N(t)-l}) f_X(t - t_{N(t)}). \tag{7}$$

We refer to this class of models as copula MTDPPs. Their conditional intensity in (5) involves hazard function components $h_l(u \,|\, v) = f_l(u \,|\, v)/S_l(u \,|\, v)$, where $S_l(u \,|\, v) = 1 - \partial C_l(F_{U_l}(u), F_{V_l}(v))/\partial F_{V_l}$. A closed-form expression for h_l relies on the specific copula function (e.g., a Gaussian copula leads to an analytically intractable h_l).

For certain copulas, the conditional and marginal densities belong to the same family of distributions. In particular, consider the three-parameter Burr density, $\text{Burr}(x \,|\, \gamma, \lambda, \psi) = \psi\gamma x^{\gamma-1}\lambda^{-\gamma}\{1 + (x/\lambda)^{\gamma}\}^{-(\psi+1)}$, with shape parameters $\gamma > 0$, $\psi > 0$, and scale parameter $\lambda > 0$. The associated hazard function is monotonically decreasing when $0 < \gamma \leq 1$, and hump-shaped when $\gamma > 1$. In the supplementary material, we provide details for the derivation of a bivariate Burr distribution built from Burr marginals and a heavy right tail copula, such that the conditionals are also Burr distributions.

To construct a class of Burr MTDPPs, for each l, we specify f_{U_l,V_l} with the bivariate Burr density such that the marginals are $f_{U_l}(x) = f_{V_l}(x) = f_X(x) = \text{Burr}(x \mid \gamma, \lambda, \kappa - 1)$, where $\kappa > 1$. Then, the conditional density, $f_{U_l|V_l}(u \mid v) = \text{Burr}(u \mid \gamma, \tilde{\lambda}(v), \kappa)$, where $\tilde{\lambda}(v) = (\lambda^\gamma + v^\gamma)^{1/\gamma}$. Hence, the conditional duration density of the Burr MTDPP is

$$f^*(t - t_{N(t)}) = \sum_{l=1}^{L} w_l \, \text{Burr}(t - t_{N(t)} \mid \gamma, \tilde{\lambda}(t_{N(t)-l+1} - t_{N(t)-l}), \kappa), \tag{8}$$

with stationary marginal $f_X(t - t_{N(t)}) = \text{Burr}(t - t_{N(t)} \mid \gamma, \lambda, \kappa - 1)$.

In fact, the stationary Burr MTDPP model includes special cases. If $\gamma = 1$, it becomes a Lomax MTDPP with stationary marginal $P(t - t_{N(t)} \mid \lambda, \kappa - 1)$. Moreover, when $\kappa = 2$, it reduces to a model with stationary log-logistic marginal $LL(t - t_{N(t)}) \mid \gamma, \lambda)$, where $LL(x \mid \gamma, \lambda) = \gamma x^{\gamma-1} \lambda^{-\gamma} \{1 + (x/\lambda)^\gamma\}^{-2}$.

2.4 Extension to MTD cluster point processes

A self-exciting MTDPP encourages clustering behavior for the arrival times. In practice, there may exist different factors that drive the duration process dynamics. As an example form hydrology, durations of dry spells can be classified into two types, corresponding to cyclonic and anticyclonic weather (Cowpertwait, 2001). A point process model for such data should be able to account for the two weather types, as the lengths of the dry spells may be distinctly different. Similar examples can also be found in Li et al. (2021). Here, we extend MTDPPs to MTD cluster point processes (MTDCPPs), based on a two-component mixture model. The definition is given as follows.

Definition 2. Let $N(t)$ be a temporal point process defined on $\mathbb{R}^+$ with event arrival times $t_1, t_2, \cdots \in \mathbb{R}^+$. Let $f^*(t - t_{N(t)})$ be the conditional duration density of a self-exciting MTD point process. Then, $N(t)$ is said to be an *MTD cluster point process* if (i) $t \sim f_I$ for

$N(t) = 0$; (ii) for $N(t) \geq 1$, the conditional duration density is given by

$$f_C^*(t - t_{N(t)}) \; = \; \pi_0 \, f_I(t - t_{N(t)}) + (1 - \pi_0) \, f^*(t - t_{N(t)}), \tag{9}$$

where $0 \leq \pi_0 \leq 1$, and f_I is a density on $\mathbb{R}^+$.

Similar to the MTDPP, we use densities f_I and f_C^* to define the conditional arrival densities p_i^* of event time t_i, for an observed point pattern $\{t_i\}_{i=1}^n$, by taking $p_1^*(t) = f_I(t)$ and $p_i^*(t) = f_C^*(t - t_{i-1})$, $t > t_{i-1}$, for $i = 2, \ldots, n$. When $\pi_0 = 1$, the MTDCPP reduces to a renewal process; furthermore, if f_I corresponds to the exponential distribution, it becomes a Poisson process. When $\pi_0 = 0$, the model becomes an MTDPP.

Let h_I be the hazard function associated with f_I. The conditional intensity of the MTDCPP extends the mixture form in (5) as follows:

$$\lambda_C^*(t) = \pi_0(t) \, h_I(t - t_{N(t)}) \; + \; \sum_{l=1}^{L} \pi_l(t) \, h_l(t - t_{N(t)} \, | \, t_{N(t)-l+1} - t_{N(t)-l}), \tag{10}$$

where $\pi_0(t) \; = \; \pi_0 \, S_I(t - t_{N(t)})/S_C^*(t - t_{N(t)})$, $\;\pi_l(t) \; = \; (1 - \pi_0) \, w_l \, S_l(t - t_{N(t)} \, | \, t_{N(t)-l+1} - t_{N(t)-l})/S_C^*(t - t_{N(t)})$, for $l = 1, \ldots, L$, $S_C^*(t - t_{N(t)}) = \pi_0 S_I(t - t_{N(t)}) + (1 - \pi_0) S^*(t - t_{N(t)})$, and we have that $\pi_l(t) \geq 0$, for $l = 0, \ldots, L$, and $\sum_{l=0}^{L} \pi_l(t) = 1$, for all t.

Compared to the MTDPP conditional intensity function, the MTDCPP conditional intensity has an extra term contributed by component f_I, with appropriately renormalized time-dependent weights. If we take an exponential density with rate parameter μ for f_I, and a stationary Lomax MTDPP for f^*, we refer to the resulting model as the Lomax MTDCPP. Note that we consider the stationary Lomax instead of the stationary scaled-Lomax MTDPP to avoid potential identifiability issues, indicated by Proposition 4.

3 Bayesian implementation

3.1 Conditional likelihood and prior specification

Let $\{t_i\}_{i=1}^n$ be the observed temporal point pattern over the interval $(0, T)$, with durations $x_1 = t_1$ and $x_i = t_i - t_{i-1}$, for $i = 2, \ldots, n$. We outline the approach to posterior inference for the MTDCPP model based on a conditional likelihood (the MTDPP is a special case with $\pi_0 = 0$). Further details for the specific MTDPP and MTDCPP models illustrated in Section 4 are provided in the supplementary material.

The point process likelihood can be expressed equivalently using $\{t_i\}$ or $\{x_i\}$. For brevity, we use the latter, and, for notational convenience, let $x_{n+1} = T - t_n$. Thus, combining (2), (4), and (9), the likelihood conditional on $(x_1, \ldots, x_L)$ is

$$
\begin{aligned}
p(x_1, \ldots, x_{n+1}\,;\, \pi_0, \boldsymbol{w}, \boldsymbol{\phi}, \boldsymbol{\theta}) \propto & \prod_{i=L+1}^{n} \left\{ \pi_0 f_I(x_i \,|\, \boldsymbol{\phi}) + (1-\pi_0) \sum_{l=1}^{L} w_l f_l(x_i \,|\, x_{i-l}, \boldsymbol{\theta}_l) \right\} \\
& \times \left(1 - \int_0^{x_{n+1}} \left\{ \pi_0 f_I(u \,|\, \boldsymbol{\phi}) + (1-\pi_0) \sum_{l=1}^{L} w_l f_l(u \,|\, x_{n+1-l}, \boldsymbol{\theta}_l) \right\} du \right)
\end{aligned} \tag{11}
$$

where $\boldsymbol{w} = (w_1, \ldots, w_L)^\top$, and the vectors $\boldsymbol{\phi}$ and $\boldsymbol{\theta} = \{\boldsymbol{\theta}_l\}_{l=1}^L$ collect the parameters of the independent duration density f_I and the MTDPP component densities, respectively.

A Bayesian model formulation involves priors for the probability π_0, the MTDPP weight vector $\boldsymbol{w}$, as well as density parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$. The priors for $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ depend on particular choices of the densities f_I and f_l. For π_0, we consider a beta prior $\text{Beta}(\pi_0 \,|\, u_0, v_0)$ (without further information, we take the uniform prior, $\text{Beta}(\pi_0 \,|\, 1, 1)$).

We take the weights w_l as increments of a c.d.f. G, i.e., $w_l = G(l/L) - G((l-1)/L)$, for $l = 1, \ldots, L$, where G has support on the unit interval. Flexible estimation of the weights depends on the shape of G. Thus, we consider a Dirichlet process (DP) prior (Ferguson, 1973) which supports general distributional shapes for G, denoted as $\text{DP}(\alpha_0, G_0)$, where $G_0 = \text{Beta}(a_0, b_0)$ is the baseline c.d.f., and $\alpha_0 > 0$ is the precision parameter. Given G_0

and α_0, the vector of weights $\boldsymbol{w}$ follows a Dirichlet distribution with shape parameter vector $\alpha_0(a_1, \dots, a_L)^\top$, where $a_l = G_0(l/L) - G_0((l-1)/L)$, for $l = 1, \dots, L$. The prior expectation is $E(\boldsymbol{w}) = (a_1, \dots, a_L)^\top$. We denote this prior for the weights as $\text{CDP}(\cdot \mid \alpha_0, a_0, b_0)$. As it is natural to assume that near lagged durations contribute more than distant ones, our default choice for G_0 is $\text{Beta}(1, b_0)$, with $b_0 > 1$, such that the weights exhibit a decreasing pattern in prior expectation. The DP-based prior for the weights supports the strategy of fitting an over-specified model, i.e., with a large value of L. Model inference induces the contribution of the lagged durations, assigning large weights to important ones. In practice, the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the observed duration time series can also be used to guide the choice of L.

3.2 Posterior simulation

We outline an MCMC posterior simulation method for the model parameters. For more efficient notation, we rewrite the MTDCPP transition density as $f_C^*(x_i) = \sum_{l=0}^{L} \pi_l f_l^c(x_i \mid \boldsymbol{\phi}, \boldsymbol{\theta}_l)$, where $f_0^c = f_I$, $f_l^c = f_l$, $\pi_l = (1 - \pi_0) w_l$, for $l = 1, \dots, L$, and $\sum_{l=0}^{L} \pi_l = 1$.

We augment the model with configuration variables ℓ_i, taking values in $\{0, 1, \dots, L\}$, with discrete distribution $\sum_{l=0}^{L} \pi_l \, \delta_l(\ell_i)$, where $\delta_l(\ell_i) = 1$ if $\ell_i = l$ and 0 otherwise, for $i = L+1, \dots, n$. Therefore, $\ell_i = 0$ indicates that the duration x_i is generated from f_I, and $\ell_i = l$ indicates that x_i is generated from the lth component of the MTDPP, for $l = 1, \dots, L$. Note that the likelihood normalizing term in (11) can be written as $\sum_{l=0}^{L} \pi_l \, S_l^c(x_{n+1} \mid \boldsymbol{\phi}, \boldsymbol{\theta}_l)$, where $S_0^c = S_I$ and $S_l^c = S_l$, for $l = 1, \dots, L$. Similarly with the observed durations, we can introduce a configuration variable ℓ_{n+1} to identify the component of the mixture for x_{n+1}.

The posterior distribution of the augmented model is proportional to

$$p(\boldsymbol{\phi}) \times \prod_{l=1}^{L} p(\boldsymbol{\theta}_l) \times \mathrm{Dir}(\boldsymbol{w} \,|\, \alpha_0 a_1, \ldots, \alpha_0 a_L) \times \mathrm{Beta}(\pi_0 \,|\, u_0, v_0)$$
$$\times \left\{ \prod_{i=L+1}^{n} f^c_{\ell_i}(x_i \,|\, \boldsymbol{\phi}, \boldsymbol{\theta}_l) \sum_{l=0}^{L} \pi_l \, \delta_l(\ell_i) \right\} \left\{ S^c_{\ell_{n+1}}(x_{n+1} \,|\, \boldsymbol{\phi}, \boldsymbol{\theta}_l) \sum_{l=0}^{L} \pi_l \, \delta_l(\ell_{n+1}) \right\}$$

Posterior updates for parameters $\boldsymbol{\phi}$ and $\{\boldsymbol{\theta}_l\}$ depend on f_I and f_l, respectively. Implementation details for the models of Section 4 are provided in the supplementary material. The posterior full conditional distribution of ℓ_i is a discrete distribution on $\{0, ..., L\}$ with probabilities proportional to $\pi_l f^c_l(x_i \,|\, \boldsymbol{\phi}, \boldsymbol{\theta}_l)$, for $i = L+1, \ldots, n$, and with probabilities proportional to $\pi_l S^c_l(x_{n+1} \,|\, \boldsymbol{\phi}, \boldsymbol{\theta}_l)$, for $i = n+1$. Let $M_l = |\{i : \ell_i = l\}|$, for $l = 0, ..., L$, where $|\{\cdot\}|$ returns the size of set $\{\cdot\}$. Given the configuration variables, we update the weights $\boldsymbol{w}$ with a Dirichlet posterior full conditional distribution with parameter vector $(\alpha_0 a_1 + M_1, \ldots, \alpha_0 a_L + M_L)^\top$. The beta prior for π_0 yields a conjugate posterior full conditional distribution, $\mathrm{Beta}(\pi_0 \,|\, u_0 + M_0, v_0 + \sum_{l=1}^{L} M_l)$.

3.3 Inference for point process functionals and model checking

Using the MCMC algorithm, we obtain posterior samples that provide full inference for any functional of the point process. For example, given the posterior draws for the model parameters, we obtain posterior realizations for the conditional intensity function by evaluating (5) or (10) over a grid of time points. Similarly, for stationary MTDPPs, we can obtain point and interval estimates for the marginal duration density.

For model assessment, we use the time-rescaling theorem (Daley and Vere-Jones, 2003), according to which $\{\Lambda^*(t_i)\}_{i=1}^n$ is a realization from a unit rate Poisson process, where $\Lambda^*(t) = \int_0^t \lambda^*(u) du$ is the conditional cumulative intensity, and $\{t_1 < t_2 < ... < t_n\}$ is the observed point pattern. Hence, if the model is correctly specified, $U^*_i = 1 - \exp\{-(\Lambda^*(t_i) -$

$\Lambda^*(t_{i-1}))\}$, for $i = 1, \ldots, n$, are independent uniform random variables on $(0, 1)$. Thus, the model can be assessed graphically using quantile-quantile plots for the estimated U_i^*.

For MTDPP models, $\Lambda^*(t_i) = \sum_{j=1}^{i} \int_{t_{j-1}}^{t_j} h^*(u - t_{j-1})du$, and thus $\Lambda^*(t_i) - \Lambda^*(t_{i-1}) = \int_{t_{i-1}}^{t_i} h^*(u - t_{i-1})du$. Using the relationship between the conditional survival and cumultive intensity functions, we have $S^*(t - t_{i-1}) = \exp(-\int_{t_{i-1}}^{t} h^*(u - t_{i-1})du)$, for $t_{i-1} < t \leq t_i$. Therefore, $S^*(t_i - t_{i-1}) = \exp\{-(\Lambda^*(t_i) - \Lambda^*(t_{i-1}))\}$, which allows us to obtain posterior samples for the U_i^* from $U_i^* = 1 - S^*(t_i - t_{i-1}) = 1 - \sum_{l=1}^{L} w_l \, S_l(t_i - t_{i-1} \,|\, t_{i-l} - t_{i-l-1}, \boldsymbol{\theta}_l)$. Replacing survival function S^* with S_C^*, the approach can also be used for MTDCPPs.

4 Data illustrations

We illustrate the scope of the modeling framework through one synthetic and two real data examples. In the simulation example, we explore inference results for conditional intensities and duration hazard functions of different shapes. The goal of the first real data example is to demonstrate the practical utility of stationary MTDPPs for scenarios where the independence assumption of renewal processes needs to be relaxed. The second real data example examines the capacity of MTDCPPs to detect and quantify duration clustering behaviors; this was also evaluated through a simulation study, the details of which can be found in the supplementary material. Also available in the supplementary material are graphical model assessment results for all data examples, obtained using the approach of Section 3.3; the results indicate good model fit for all data examples.

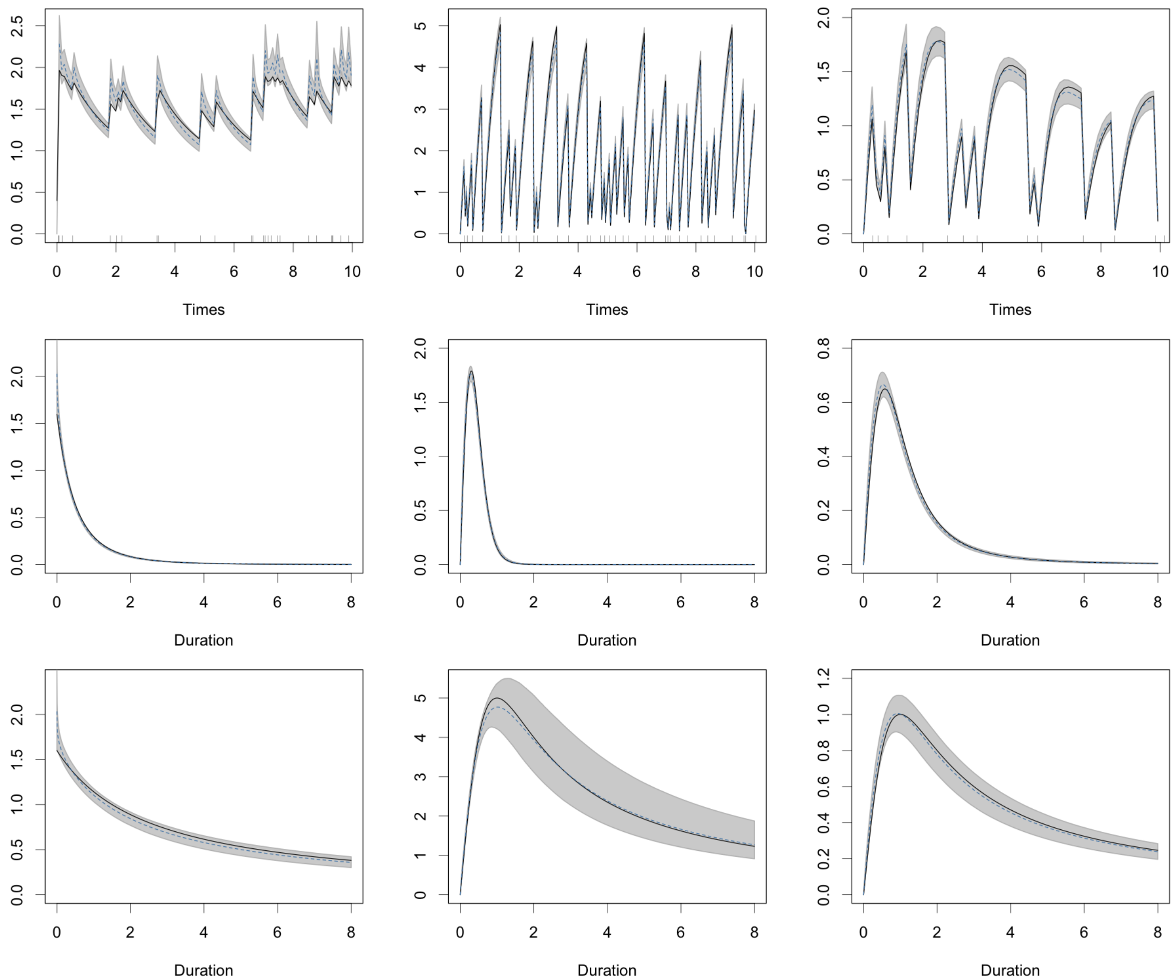
The results for each example are based on posterior samples collected every fourth iteration from a Markov chain of 25000 iterations with a burn-in of 5000 samples. We implemented all algorithms in the R programming language, with C++ code integrated to update latent variables, on a computer with a 2-GHz Intel Core i5 processor and 32-GB RAM. The computing time was about 2 minutes and 20 minutes for the synthetic data and for the first real data example, respectively. The second real data example involves 121 point patterns; the average computing time for a point pattern was around 5 minutes.

4.1 Simulation experiment

We generated data from three stationary MTDPP models (discussed in Section 2.3) with scaled-Lomax, Burr, and log-logistic marginal marginal duration distributions. The respective parameters were set at $(\phi, \alpha) = (0.5, 5)$, $(\lambda, \gamma, \kappa) = (1, 2, 6)$, and $(\lambda, \gamma) = (1, 2)$, such that the hazard function for the durations is decreasing for the scaled-Lomax MTDPP, and hump-shaped for the other two models; see Figure 1. The model order was $L = 3$ for all simulations, with decaying weights $\boldsymbol{w} = (0.5, 0.3, 0.2)$. For each simulated point pattern, we chose the observation window to obtain around 2000 event times.

We applied the Burr MTDPP model in (8), with $L = 3$, to the three synthetic data sets. Recall that the hazard function of the marginal $\text{Burr}(\gamma, \lambda, \kappa - 1)$ duration distribution is decreasing when $\gamma \leq 1$, and hump-shaped when $\gamma > 1$. We thus assigned a $\text{Ga}(1, 1)$ prior to γ, where $\text{Ga}(a, b)$ denotes the gamma distribution with mean a/b. Moreover, the mth moment of the $\text{Burr}(\gamma, \lambda, \kappa - 1)$ distribution exists if $\gamma(\kappa - 1) > m$. Independently of γ, we placed a truncated gamma prior, $\text{Ga}(6, 1)\mathbb{1}(\kappa > 1)$, on κ. Since $\text{E}(\kappa) = 6.004$, the prior choice for γ and κ implies that, in prior expectation, the first five moments of the marginal duration distribution exist. The scale parameter λ was assigned a $\text{Ga}(1, 1)$ prior, and the vector of weights a $\text{CDP}(5, 1, 2)$ prior.

Figure 1 plots point and interval estimates for the point process conditional intensity, as well as for the duration process marginal density and its associated hazard function. Note that, although the true data generating mechanisms correspond to MTDPPs, the Burr MTDPP is a mis-specified model for two of the simulated data sets. However, the model is able to distinguish between monotonically decreasing and hump-shaped shapes for the hazard function associated with the marginal duration distribution. Overall, based on a



(a) scaled-Lomax (b) Burr (c) Log-logistic

Figure 1: Synthetic data example. The first, second, and third rows plot posterior means (blue dashed lines) and 95% credible interval estimates (grey bands) for the conditional intensity, marginal duration density, and marginal duration hazard function. The black solid lines correspond to the true functions.

single process realization, the Burr MTDPP model provided reasonably accurate estimates for different point process functionals, with uncertainty bands that effectively contain the true functions.

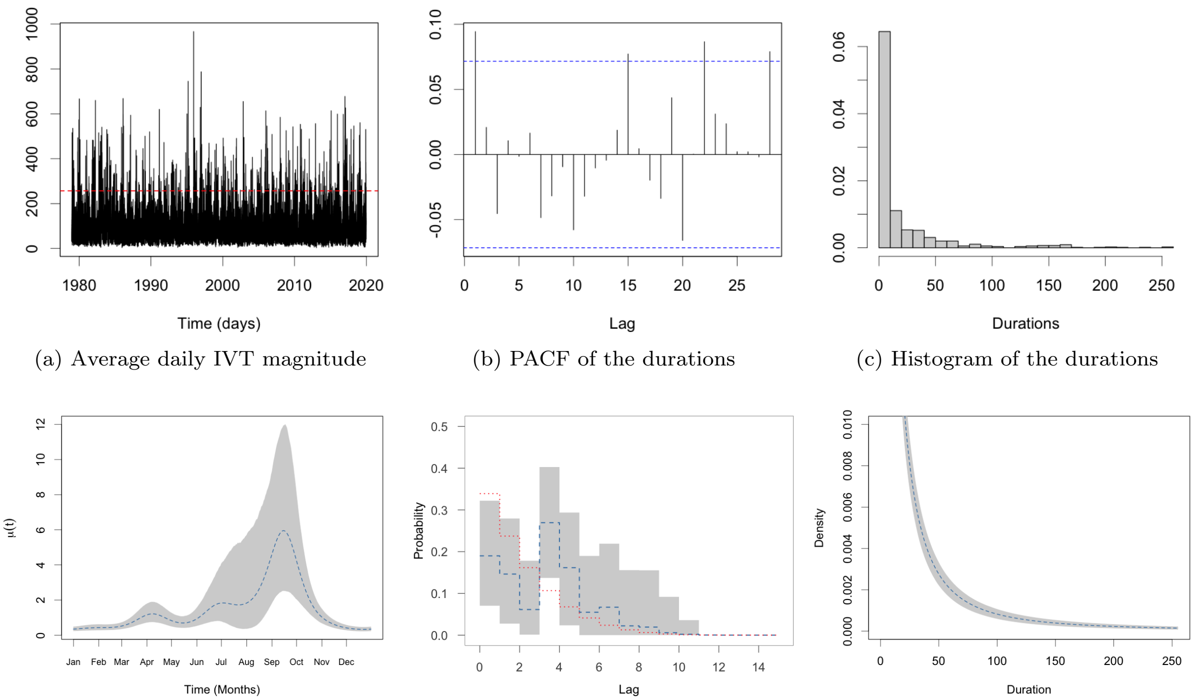
4.2 IVT recurrence interval analysis

Integrated water vapor transport (IVT) is a vector representing the total amount of water vapor being transported in an atmospheric column. Atmospheric rivers (ARs), which are corridors of enhanced IVT, play a vital role in transporting moisture into western North

America. Identifying and tracking ARs is central to understanding high-impact weather events, such as extreme precipitation and flooding (Hughes et al., 2022). Rutz et al. (2019) review several of the AR detection algorithms, most of which use IVT thresholds as input. Typically, an IVT threshold is chosen based on a specified quantile. Appropriately thresholding the IVT is important to improve AR detection; e.g., Barata et al. (2022) provide a time-varying quantile estimate of the IVT using a dynamic statistical model.

In this example, we take on a different perspective to study the IVT, based on the general idea that strong ARs tend to associate with extreme IVT magnitudes. We obtain a collection of recurrent events for which the IVT magnitude exceeds a given threshold. The collection is considered as a realization of a point process. Modeling extreme events using a point process approach is motivated by the asymptotic behavior of threshold exceedances. This approach assumes that, for a large threshold, the exccedances and the associated event times can be considered as a marked Poisson process; see, e.g., Kottas and Sansó (2007) and further references therein. On the other hand, the Poisson process assumption may be too restrictive, as well as unsuitable for applications where the inferential interest lies in the stationary distribution of the durations between event times. Studying the recurrence interval distribution is important in many areas, including study of earthquakes above a certain magnitude (Corral, 2004), extreme returns (Jiang et al., 2018), and large volatilities (Jiang et al., 2016). Depending on the correlation structure of the original time series, the recurrence interval distribution may exhibit different types of tail behavior, such as power law. Furthermore, the recurrence intervals themselves can be dependent (Santhanam and Kantz, 2008). In this case, a generalization of the renewal process is needed in order to capture the dependence among durations.

Here, we demonstrate the potential of MTDPPs for the aforementioned goal, that is, simultaneous modeling the stationary recurrence interval distribution and capturing the intervals dependence. The data set involves a time series of average daily IVT magnitude



(a) Average daily IVT magnitude (b) PACF of the durations (c) Histogram of the durations

(d) Harmonic function $\mu(t)$ (e) Estimated weights (f) Stationary marginal estimate

Figure 2: Recurrence interval analysis. Panel (a) shows the average daily IVT magnitude with the 0.95 quantile (red line) fixed over time. Panels (b) and (c) plot the PACF and histogram of the recurrence intervals, respectively. Panel (d) shows the harmonic function $\mu(t)$ for an one-year window. Panel (e) shows inference results for the weights. Panel (f) is the close view of the stationary marginal estimate, specifically at the tail of the marginal density. In Panels (d), (e), and (f), the blue dashed line and the light grey polygon correspond to the posterior mean and the pointwise 95% credible interval estimates, respectively. The red dotted line in panel (e) is the prior mean.

calculated using ERA5, a climate reanalysis that provides hourly estimates of atmospheric variables. The time series, shown in Figure 2(a), has 14965 observations, spanning from January 1, 1979 to December 31, 2019, with all February 29s omitted, corresponding to the Santa Cruz city in California. The data are publicly available in the R package *exdqlm*. Using the 0.95 quantile threshold, we obtained 749 point events of IVT exceedances. The PACF and histogram of the durations in Figure 2(b)-(c) suggest dependence in the durations and a heavy right tail for the recurrence interval distribution.

We consider the scaled-Lomax MTDPP model. As previously discussed in Section 2.3, the model has a stationary scaled-Lomax marginal distribution $P(\alpha\phi, \alpha - 1)$ for the recurrence intervals, and the conditional duration distribution converges to the exponential distribution with rate parameter ϕ^{-1}, as $\alpha \to \infty$. Let $\{t_i\}$ and $\{x_i\}$ be the observed event times and durations, respectively. To account for potential seasonality, we use the following multiplicative model, $x_i = \mu(t_i)z_i$, with $\log \mu(t_i) = \sum_{j=1}^{J}\{\beta_{1j}\sin(j\omega t_i) + \beta_{2j}\cos(j\omega t_i)\}$, where $\omega = 2\pi/T_0$, and $T_0 = 365$ is the period for daily data. We assume the stationary scaled-Lomax MTDPP model for the z_i, such that the conditional duration density is $f^*(x_i) = \mu(t_i)^{-1}\sum_{l=1}^{L} w_l P\left(\mu(t_i)^{-1}x_i \,|\, \alpha\phi + \mu(t_{i-l})^{-1}x_{i-l}, \alpha\right)$.

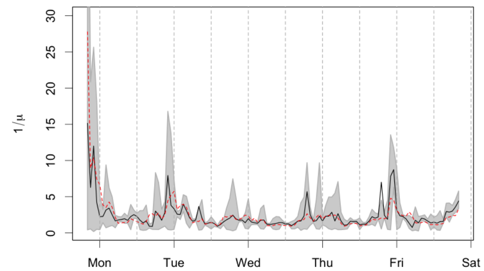
On the basis of the ACF and PACF plots, we chose MTDPP model order $L = 15$. We took $J = 5$, and assigned mean-zero, dispersed normal priors to the regression parameter vector. The shape and scale parameters α and ϕ received $\text{Ga}(6,1)\mathbb{1}(\alpha > 1)$ and $\text{Ga}(1,1)$ priors, respectively. For the weights, we consider a $\text{CDP}(w \,|\, 5,1,6)$ prior, which implies a decreasing trend in prior expectation (see Figure 2(e)).

The posterior mean and 95% credible interval estimates of the harmonic component coefficients imply the presence of annual and semiannual seasonality. The posterior estimates of the corresponding coefficients $(\beta_{11}, \beta_{21}, \beta_{22})$ are $-0.59\,(-0.86, -0.31)$, $-0.71\,(-1.08, -0.36)$, and $-0.52\,(-0.82, -0.20)$. Figure 2(d) shows the function $\mu(t)$ evaluated at a grid over a period of one year. Smaller durations between high IVT magnitudes tend to appear from November to March, corresponding to high atmospheric river frequency during that period. In fact, this time interval corresponds to the usual flooding period in California (e.g., the most recent floods in California were caused by multiple atmospheric rivers between December 2022 and March 2023). Figure 2(e) shows the estimated weights. Lags one, two, four and five are the most influential, which suggests serial dependence in the durations. The posterior mean and 95% credible interval estimates of α and ϕ were $2.01\,(1.73, 2.37)$ and $5.03\,(3.40, 7.00)$, respectively. Note that the stationary marginal distribution of the process $\{z_i\}$ is $P(z \,|\, \alpha\phi, \alpha - 1)$, with finite mean for $\alpha > 2$ and finite variance for $\alpha > 3$. Inference for α suggests that, even after adjusting for seasonality, the distribution of the recurrence intervals is heavy tailed. In fact, Figure 2(f) shows a tail for the marginal density that decays very slowly, in particular when compared to the histogram of the observed durations in Figure 2(c), where the seasonality is not accounted for. The heavy-tailed recurrence interval distribution also indicates a cluster phenomenon of the IVT extremes.

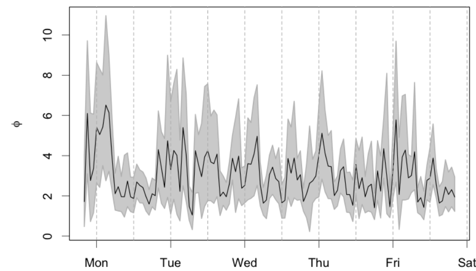
4.3 Mid-price changes of the AUD/USD exchange rate

Financial markets involve complex human activities, with both external and internal factors driving market dynamics. It is suggested that, for high-frequency financial data, only a small portion of the price movements is caused by external factors such as relevant news releases (Filimonov and Sornette, 2012). Therefore, to understand the financial market microstructure, it is important to quantify the degree of market reflexivity, measured as the proportion of price movements due to internal rather than external processes. The Hawkes process and its extensions have been used to study market reflexivity, where each price move is considered as an event (Filimonov and Sornette, 2012; Wheatley et al., 2016; Chen and Stindl, 2018). The Hawkes process admits a branching structure that allows for separating endogenous from exogenous events, including a branching ratio parameter that can be used to quantify market reflexivity. Here, we explore modeling for market reflexivity from the duration clustering perspective using the MTDCPP, where the probability $(1 - \pi_0)$ in (9) can be interpreted as the proportion of price movements due to endogenous interaction. At the end of the section, we discuss our findings relative to Hawkes process based models.

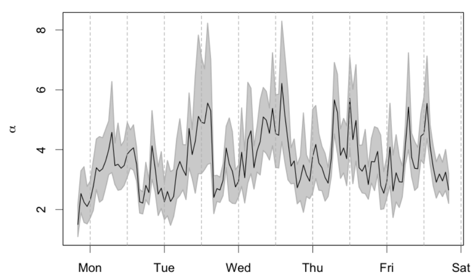
We analyze the price movements of the AUD/USD foreign exchange rate. A price movement is recorded when there is a mid-price change, where mid-price is defined as the average of the best bid and ask prices. A detailed explanation of using mid-price change as a measure of price movements can be found in Filimonov and Sornette (2012). The data consists of 121 non-overlapping point patterns, with total number of events ranging from 108 to 3961. Each point pattern corresponds to an one-hour time window of the trading week from 20:00:00 Greenwich Mean Time (GMT) July 19 to 21:00:00 GMT July 24 in year 2015. Analyzing sequences of point patterns within small time windows avoids to



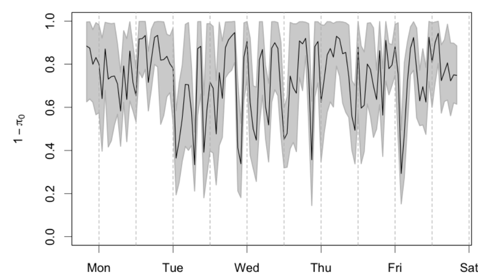
(a) Exponential distribution waiting time



(b) Lomax MTDPP scale parameter ϕ



(c) Lomax MTDPP shape parameter α



(d) Reflexivity level $(1 - \pi_0)$

Figure 3: AUD/USD foreign exchange market reflexivity analysis. Time series of the posterior means (solid lines) and pointwise 95% credible intervals (grey polygons) for parameters $1/\mu$, ϕ, α, $(1-\pi_0)$, based on the MTDCPP model. Vertical dashed lines correspond to midnight and midday GMT. The red dashed line in panel (a) corresponds to the averages of the observed durations of the one hour windows.

some extent the issue of nonstationarity, such as diurnal pattern. We refer to Chen and Stindl (2018) for more details about the data, which is available in R package *RHawkes* (Chen and Stindl, 2022).

We considered the Lomax MTDCPP model, that is, model (9) with f_I given by the exponential density with rate parameter μ, and f^* corresponding to the stationary Lomax MTDPP. We used a Beta$(1,1)$ prior for probability π_0, and a Ga$(1,1)$ prior for μ. For the stationary Lomax MTDPP, the shape and scale parameters received priors Ga$(\alpha \,|\, 6,1)\mathbb{1}(\alpha > 1)$ and Ga$(\phi \,|\, 1,1)$, respectively. Based on the autocorrelation and partial autocorrelation functions of the observed durations, we chose model order $L = 15$ for all point patterns, and the mixture weights were assigned a CDP$(\boldsymbol{w} \,|\, 5,1,6)$ prior, which elicits a decreasing pattern in the weights.

We applied the model to each of the 121 point patterns. Figure 3 shows the time series of posterior mean and interval estimates for four parameters: exponential distribution parameter $1/\mu$, Lomax MTDPP scale and shape parameters ϕ and α, and the endogenous probability $(1 - \pi_0)$. Note that the exponential distribution and Lomax MTDPP are regarded as drivers of external and internal factors for waiting times between successive mid-price changes, respectively. The estimates of the mean waiting time $1/\mu$ for external factors shows obvious diurnal pattern, with peaks and troughs appearing around midnight and midday GMT, respectively. The posterior estimates of ϕ for all point patterns seem more volatile, with relatively high and low values occuring at midnight and midday GMT, whereas the posterior estimates of α reflect an opposite pattern. The mean of the stationary marginal distribution of the Lomax MTDPP is $\phi/(\alpha-2)$, provided that $\alpha > 2$. Thus, given the patterns of estimated ϕ and α, the estimates of the mean waiting time for internal factors appear high and low around midnight and midday GMT, respectively. In addition, small values of α around midnight GMT suggest a heavy-tailed duration distribution of the Lomax MTDPP during that period. This indicates that mid-price changes tend to cluster around midnight GMT, which corresponds to the opening time of Asian markets (23:00-1:00 GMT).

As shown in Figure 3(d), the estimates of the market reflexivity $(1 - \pi_0)$ fluctuate heavily over the whole trading week, with most of them greater than 0.5. The posterior means of $(1 - \pi_0)$ for the 121 point patterns range from 0.29 to 0.95 with median 0.76 and quartiles $(0.64, 0.87)$, suggesting that the market dynamics are mostly driven by internal processes. A similar conclusion was drawn by Chen and Stindl (2018) where a renewal Hawkes (RHawkes) process was used for the same data. The RHawkes process (Wheatley et al., 2016) extends the Hawkes process to capture dependence between clusters, by replacing the immigrant Poisson process with a renewal process. Both the Hawkes and RHawkes process models include the branching ratio parameter which quantifies market reflexivity. The branching ratio parameter estimates reported by Chen and Stindl (2018) range from 0.29 to 0.98 with median 0.66 and quartiles $(0.53, 0.80)$. Under a different stochastic model structure, the Lomax MTDCPP is able to quantify the extent to which the observed dynamics are caused by internal factors versus external influences.

Finally, we note that using the MTDCPP for the present example does not require any stationarity assumptions. In contrast, stationarity is essential for both the Hawkes and RHawkes processes in order to use the branching ratio as an estimator for the market reflexivity. However, as discussed in Filimonov and Sornette (2012), market activities are typically nonstationary. The lack of stationarity is typically attributed to seasonal trends, which can be addressed by splitting the time window into small intervals, as shown in this example. Still, one has to balance the size of the intervals and the number of the events within the interval to ensure reliable estimates are produced. Moreover, even after removing seasonality, stationarity is not necessarily guaranteed. Therefore, MTDCPP models may be useful in applications where stationarity assumptions are not plausible.

5 Summary and discussion

We have developed a new class of stochastic models for temporal point patterns with self-excitation or self-regulation effects, identically distributed but dependent durations, or clustered durations. The modeling framework allows for different approaches to building the point process: through marginal duration distributions, when the inferential goal pertains to the intervals between event times; or through conditional hazard functions, when interest lies in the point process dependence structure on its history. Both strategies connect naturally to existing point process models. The former is analogous to renewal process modeling, while the latter involves the same motivation with Hawkes processes. Our main objective is to provide a new modeling and inference tool for a range of applications involving temporal point patterns. To the best of our knowledge, the proposed class of point processes is the first one that enables simultaneous modeling of high-order dependence and stationary durations, with computationally efficient inference.

Our framework builds from a structured mixture model for the point process conditional duration density. The resulting point process has restricted memory, i.e., its evolution depends on recent events. This assumption is generally suitable for relatively large point patterns. For scenarios where one anticipates more extensive history dependence, a large value for the order of the mixture model can be used. The nonparametric prior for the weights allows efficient inference with a large order. On the other hand, there are applications where data correspond to many processes that exhibit a relatively small number of point events, such as the analysis of recurrent event gap times for multiple patients in medical studies. For such data, a small order is more appropriate. In fact, even the special case where the conditional duration density depends on the most recent lag provides a meaningful generalization of renewal processes commonly used for this type of analysis.

In many applications, point patterns include information on marks, that is, random variables associated with each point event, such that the data generating mechanism corresponds to a marked point process. Consider, for instance, continuous marks, $\boldsymbol{y}$. The marked point process intensity can be developed from $\lambda^*(t, \boldsymbol{y}) = \lambda_g^*(t)\, m_t^*(\boldsymbol{y})$, where $\lambda_g^*(t)$ is the conditional intensity for the event times (referred to as the ground process intensity), and $m_t^*(\boldsymbol{y})$ is the time-dependent mark distribution (Daley and Vere-Jones, 2003). The proposed framework can be utilized for marked point processes by combining an MTDPP or MTDCPP model for the ground process with a model for the mark distribution.

Supplementary material

The supplementary material includes proofs for the theoretical results, details for the bivariate Burr distribution, MCMC implementation details, an additional simulation experiment, and model checking results.

References

Arnold, B. C., Castillo, E., and Sarabia, J. M. (1999), *Conditional Specification of Statistical Models*, New York: Springer.

Barata, R., Prado, R., and Sansó, B. (2022), "Fast inference for time-varying quantiles via flexible dynamic models with application to the characterization of atmospheric rivers," *The Annals of Applied Statistics*, 16, 247–271.

Bhogal, S. K. and Thekke Variyam, R. (2019), "Conditional duration models for high-frequency data: a review on recent developments," *Journal of Economic Surveys*, 33, 252–273.

Chen, F. and Stindl, T. (2018), "Direct likelihood evaluation for the renewal Hawkes process," *Journal of Computational and Graphical Statistics*, 27, 119–131.

— (2022), *RHawkes: Renewal Hawkes Process*. R package version 1.0.

Coen, A., Gutiérrez, L., and Mena, R. H. (2019), "Modelling failures times with dependent renewal type models via exchangeability," *Statistics*, 53, 1112–1130.

Cook, R. J., Lawless, J. F., et al. (2007), *The Statistical Analysis of Recurrent Events*, New York: Springer.

Corral, Á. (2004), "Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes," *Physical Review Letters*, 92, 108501.

Cowpertwait, P. S. (2001), "A renewal cluster model for the inter-arrival times of rainfall events," *International Journal of Climatology*, 21, 49–61.

Daley, D. J. and Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*, New York: Springer.

Engle, R. F. and Russell, J. R. (1998), "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, 1127–1162.

Ferguson, T. S. (1973), "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, 209–230.

Filimonov, V. and Sornette, D. (2012), "Quantifying reflexivity in financial markets: Toward a prediction of flash crashes," *Physical Review E*, 85, 056108.

Gaver, D. P. and Lewis, P. (1980), "First-order autoregressive gamma sequences and point processes," *Advances in Applied Probability*, 12, 727–745.

Grammig, J. and Maurer, K.-O. (2000), "Non-monotonic hazard functions and the autoregressive conditional duration model," *The Econometrics Journal*, 3, 16–38.

Hassan, M. Y. and El-Bassiouni, M. Y. (2013), "Modelling Poisson marked point processes using bivariate mixture transition distributions," *Journal of Statistical Computation and Simulation*, 83, 1440–1452.

Hassan, M. Y. and Lii, K.-S. (2006), "Modeling marked point processes via bivariate mixture transition distribution models," *Journal of the American Statistical Association*, 101, 1241–1252.

Hautsch, N. (2011), *Econometrics of Financial High-Frequency Data*, New York: Springer Science & Business Media.

Hawkes, A. G. (1971), “Point spectra of some mutually exciting point processes,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 33, 438–443.

Hughes, M., Swales, D., Scott, J. D., Alexander, M., Mahoney, K., McCrary, R. R., Cifelli, R., and Bukovsky, M. (2022), “Changes in extreme integrated water vapor transport on the US west coast in NA-CORDEX, and relationship to mountain and inland precipitation,” *Climate Dynamics*, 1–23.

Jacobs, P. and Lewis, P. (1977), “A mixed autoregressive-moving average exponential sequence and point process (EARMA 1, 1),” *Advances in Applied Probability*, 9, 87–104.

Jiang, Z.-Q., Canabarro, A., Podobnik, B., Stanley, H. E., and Zhou, W.-X. (2016), “Early warning of large volatilities based on recurrence interval analysis in Chinese stock markets,” *Quantitative Finance*, 16, 1713–1724.

Jiang, Z.-Q., Wang, G.-J., Canabarro, A., Podobnik, B., Xie, C., Stanley, H. E., and Zhou, W.-X. (2018), “Short term prediction of extreme returns based on the recurrence interval analysis,” *Quantitative Finance*, 18, 353–370.

Kottas, A. and Sansó, B. (2007), “Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis,” *Journal of Statistical Planning and Inference*, 137, 3151–3163.

Le, N. D., Martin, R. D., and Raftery, A. E. (1996), “Modeling flat stretches, bursts outliers in time series using mixture transition distribution models,” *Journal of the American Statistical Association*, 91, 1504–1515.

Li, X., Genest, C., and Jalbert, J. (2021), “A self-exciting marked point process model for drought analysis,” *Environmetrics*, 32, e2697.

Modarres, M., Kaminskiy, M. P., and Krivtsov, V. (2017), *Reliability Engineering and Risk Analysis: A Practical Guide*, Boca Raton: CRC Press.

Ogata, Y. (1988), "Statistical models for earthquake occurrences and residual analysis for point processes," *Journal of the American Statistical association*, 83, 9–27.

O'Hara, M. (1995), *Market Microstructure Theory*, London: Blackwell.

Pacurar, M. (2008), "Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature," *Journal of Economic Surveys*, 22, 711–751.

Reinhart, A. (2018), "A review of self-exciting spatio-temporal point processes and their applications," *Statistical Science*, 33, 299–318.

Rutz, J. J., Shields, C. A., Lora, J. M., Payne, A. E., Guan, B., Ullrich, P., O'brien, T., Leung, L. R., Ralph, F. M., Wehner, M., et al. (2019), "The atmospheric river tracking method intercomparison project (ARTMIP): quantifying uncertainties in atmospheric river climatology," *Journal of Geophysical Research: Atmospheres*, 124, 13777–13802.

Santhanam, M. and Kantz, H. (2008), "Return interval distribution of extreme events and long-term memory," *Physical Review E*, 78, 051113.

Sklar, M. (1959), "Fonctions de repartition an dimensions et leurs marges," *Publications de l'Institut de Statistique de L'Université de Paris*, 8, 229–231.

Strickland, C. M., Forbes, C. S., and Martin, G. M. (2006), "Bayesian analysis of the stochastic conditional duration model," *Computational Statistics & Data Analysis*, 50, 2247–2267.

Tang, X. and Li, L. (2021), "Multivariate temporal point process regression," *Journal of the American Statistical Association*, 1–16.

Veen, A. and Schoenberg, F. P. (2008), "Estimation of space–time branching process models in seismology using an EM–type algorithm," *Journal of the American Statistical Association*, 103, 614–624.

Wheatley, S., Filimonov, V., and Sornette, D. (2016), "The Hawkes process with renewal immigration & its estimation with an EM algorithm," *Computational Statistics & Data Analysis*, 94, 120–135.

Wold, H. (1948), "On stationary point processes and Markov chains," *Scandinavian Actuarial Journal*, 1948, 229–240.

Yang, C., Delcher, C., Shenkman, E., and Ranka, S. (2018), "Clustering inter-arrival time of health care encounters for high utilizers," in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, IEEE.

Zheng, X., Kottas, A., and Sansó, B. (2022), "On construction and estimation of stationary mixture transition distribution models," *Journal of Computational and Graphical Statistics*, 31, 283–293.

Supplementary Material for "Mixture Modeling for Temporal Point Processes with Memory"

A Proofs

Proof of Theorem 1. Consider a stationary MTD point process, that is, the corresponding duration process $\{X_i : i \geq 1\}$ has a stationary marginal distribution. Thus, the durations $X_1, \dots, X_{N(t)}$ are a collection of dependent random variables that are identically distributed. We assume that the first and second moments with respect to the stationary marginal distribution exist and are finite.

Denote by $E(X_i) = \mu$ for all i. Let $T_{N(t)} = \sum_{i=1}^{N(t)} X_i$ be the last arrival time prior to time t. We have $T_{N(t)} < t < T_{N(t)+1}$, and $\frac{T_{N(t)}}{N(t)} < \frac{t}{N(t)} < \frac{T_{N(t)+1}}{N(t)}$, for $N(t) \geq 1$. Note that $T_{N(t)}/N(t)$ is the average of the durations $X_1, \dots, X_{N(t)}$. By the strong law of large numbers for dependent non-negative random variables (Korchevsky and Petrov 2010), we have that, as $t \to \infty$, $T_{N(t)}/N(t) \to \mu$ *a.s.*, since as $t \to \infty$, $N(t) \to \infty$. Observing that $T_{N(t)+1}/N(t) = \{T_{N(t)+1}/(N(t)+1)\}\{(N(t)+1)/N(t)\}$, where the first term $T_{N(t)+1}/(N(t)+1) \to \mu$ *a.s.*, and the second term $(N(t)+1)/N(t) \to 1$, we can conclude that $N(t)/t \to 1/\mu$ *a.s.*. □

Proof of Proposition 2. The definition of the conditional intensity function $\lambda^*(t)$ yields that $m(t) = E[N(t)] = E[\int_0^t \lambda^*(u)du]$, where the expectation is taken with respect to the probability distribution of the point process. Since our interest is in $\lim_{t\to\infty} m(t)/t$, consider time t large enough such that $N(t) \geq L$.

Recall that $\lambda^*(t) \equiv h^*(t - t_{N(t)}) = f^*(t - t_{N(t)})/S^*(t - t_{N(t)})$, where $h^*(t - t_{N(t)})$ and $S^*(t - t_{N(t)})$ are the hazard and survival functions, respectively, associated with $f^*(t - t_{N(t)})$. Let $t_0 = 0$. We have that

$$\begin{aligned}\int_0^t \lambda^*(u)du &= \sum_{i=1}^{N(t)} \int_{t_{i-1}}^{t_i} h^*(u - t_{i-1})du + \int_{t_{N(t)}}^{t} h^*(u - t_{N(t)})du \\ &= \sum_{i=1}^{N(t)} (-\log\{S^*(t_i - t_{i-1})\}) - \log\{S^*(t - t_{N(t)})\}.\end{aligned} \tag{1}$$

For $i = 1, \ldots, N(t)$, by Jensen's inequality, we have that

$$\begin{aligned} -\log\{S^*(t_i - t_{i-1})\} &= -\log\left\{\sum_{l=1}^{L} w_l S_l(t_i - t_{i-1} \,|\, t_{i-l} - t_{i-1-l})\right\} \\ &\leq \sum_{l=1}^{L} w_l \left(-\log\{S_l(t_i - t_{i-1} \,|\, t_{i-l} - t_{i-1-l})\}\right) \\ &= \sum_{l=1}^{L} w_l \int_{t_{i-1}}^{t_i} h_l(u - t_{i-1} \,|\, t_{i-l} - t_{i-1-l}) du \\ &= \tilde{\Lambda}^*(t_i - t_{i-1}), \end{aligned} \tag{2}$$

where $\tilde{\Lambda}^*(a - t_k) = \sum_{l=1}^{L} w_l \int_{t_k}^{a} h_l(u - t_k \,|\, t_{k-l+1} - t_{k-l}) du$. Similarly, applying Jensen's inequality, we obtain $-\log\{S^*(t - t_{N(t)})\} \leq \tilde{\Lambda}^*(t - t_{N(t)})$, and combining (1) and (2), we have that

$$\int_0^t \lambda^*(u) du \leq \sum_{i=1}^{N(t)} \tilde{\Lambda}^*(t_i - t_{i-1}) + \tilde{\Lambda}^*(t - t_{N(t)}).$$

If $h_l \leq B_l$ for all l, then $\tilde{\Lambda}^*(t_i - t_{i-1}) \leq \sum_{l=1}^{L} w_l (t_i - t_{i-1}) B_l$, for $i = 1, \ldots, N(t)$, and $\tilde{\Lambda}^*(t - t_{N(t)}) \leq \sum_{l=1}^{L} w_l (t - t_{N(t)}) B_l$. Then we have that

$$\begin{aligned} \int_0^t \lambda^*(u) du &\leq \sum_{i=1}^{N(t)} \sum_{l=1}^{L} w_l (t_i - t_{i-1}) B_l + \sum_{l=1}^{L} w_l (t - t_{N(t)}) B_l \\ &= t_{N(t)} \sum_{l=1}^{L} w_l B_l + (t - t_{N(t)}) \sum_{l=1}^{L} w_l B_l = t \sum_{l=1}^{L} w_l B_l. \end{aligned}$$

Hence, the function $m(t) \leq t \sum_{l=1}^{L} w_l B_l$. It follows that $\lim_{t\to\infty} m(t)/t \leq \sum_{l=1}^{L} w_l B_l$. $\square$

Proof of Proposition 3. Let $(U, V) = (\alpha X, \alpha Y)$, where the joint density of (X, Y) is $f_{X,Y}(x, y) \propto (\lambda_0 + \lambda_1 x + \lambda_2 y)^{-(\alpha+1)}$, which corresponds to the bivariate Lomax distribution of Arnold et al. (1999). By change of variable, we obtain the joint density of (U, V), namely,

$$f_{U,V}(u, v) \propto (\lambda_0 + \lambda_1 u/\alpha + \lambda_2 v/\alpha)^{-(\alpha+1)}, \tag{3}$$

with normalizing constant $C = \int_0^\infty \int_0^\infty (\lambda_0 + \lambda_1 u/\alpha + \lambda_2 v/\alpha)^{-(\alpha+1)} du dv = \alpha \lambda_0^{-(\alpha-1)} \{(\alpha -$

$1)\lambda_1\lambda_2\}^{-1}$. It follows that the marginal density of U is

$$f_U(u) = C^{-1}\int_0^\infty \alpha^{-2}(\lambda_0 + \lambda_1 u/\alpha + \lambda_2 v/\alpha)^{-(\alpha+1)}dv$$
$$= (\alpha-1)(\lambda_0\alpha)^{-1}\lambda_1\{1 + (\lambda_0\alpha)^{-1}\lambda_1 u\}^{-\alpha}.$$

Since u and v are symmetric in (3), the marginal density $f_V(v) = (\alpha-1)(\lambda_0\alpha)^{-1}\lambda_2\{1 + (\lambda_0\alpha)^{-1}\lambda_2 v\}^{-\alpha}$. It follows that $f_{U|V}(u \mid v) = f_{U,V}(u,v)/f_V(v) = \alpha\lambda_1(\alpha\lambda_0 + \lambda_2 v)^{-1}\{1 + \lambda_1 u(\alpha\lambda_0+\lambda_2 v)^{-1}\}^{-(\alpha+1)}$. Similarly, we have that $f_{V|U}(v \mid u) = f_{U,V}(u,v)/f_U(u) = \alpha\lambda_2(\alpha\lambda_0+\lambda_1 u)^{-1}\{1 + \lambda_2 v(\alpha\lambda_0 + \lambda_1 u)^{-1}\}^{-(\alpha+1)}$. □

Proof of Proposition 4. Consider the stationary scaled-Lomax MTDPP with marginal duration density $P(\alpha\phi, \alpha-1)$. Suppose $N(t) \geq 1$. The survival function of the conditional duration distribution can be expressed as

$$S^*(t - t_{N(t)}) = \sum_{l=1}^{t_L} w_l^* \left(1 + \frac{t - t_{N(t)}}{\alpha\phi + t_{N(t)-l+1} - t_{N(t)-l}}\right)^{-\alpha},$$

where $t_L = \min\{N(t)-1, L\}$. In particular, for $N(t) \geq L$, $w_l^* = w_l$, for $l = 1, \ldots, L$. When $1 \leq N(t) < L$, $w_l^* = 1, \ldots, t_L - 1$, and $w_{t_L}^* = 1 - \sum_{r=1}^{t_L-1} w_r$. It follows that the weights w_l^* satisfy $\sum_{l=1}^{t_L} w_l^* = 1$ for $N(t) \geq 1$.

Then, for $N(t) \geq 1$, we have that

$$S^*(t - t_{N(t)}) = \sum_{l=1}^{t_L} w_l^* \left\{\left(1 + \frac{t - t_{N(t)}}{\alpha\phi + t_{N(t)-l+1} - t_{N(t)-l}}\right)^{-(\alpha\phi + t_{N(t)-l+1} - t_{N(t)-l})}\right\}^{1/\phi} \times \left(1 + \frac{t - t_{N(t)}}{\alpha\phi + t_{N(t)-l+1} - t_{N(t)-l}}\right)^{(t_{N(t)-l+1} - t_{N(t)-l})/\phi}. \tag{4}$$

As $\alpha \to \infty$, the limits of the first term and the second term in the lth mixture component of (4) are $\exp(-(t - t_{N(t)})\phi^{-1})$ and 1, respectively. More specifically, the limit of the first term is obtained by using the results that (i) $\lim_{n\to\infty}(1 + x/n)^n = \exp(x)$; (ii) $\lim_{n\to\infty} g_1(n)/g_2(n) = \lim_{n\to\infty} g_1(n)/\lim_{n\to\infty} g_2(n)$, provided that both $\lim_{n\to\infty} g_1(n)$ and $\lim_{n\to\infty} g_2(n)$ exist, and $\lim_{n\to\infty} g_2(n) \neq 0$.

Since $\sum_{l=1}^{t_L} w_l^* = 1$, it follows that, as $\alpha \to \infty$, the survival function of the conditional duration distribution converges to $\exp(-(t - t_{N(t)})\phi^{-1})$, i.e., the survival function of the exponential distribution with rate parameter ϕ^{-1}.

□

B Bivariate Burr distribution

We consider the following three-parameter Burr distribution (Tadikamalla, 1980). A random variable X is said to follow the Burr distribution, denoted as $\mathrm{Burr}(\gamma, \lambda, \psi)$, if the cumulative distribution function (c.d.f.) of X is $F(x) = 1 - (1 + (x/\lambda)^\gamma)^{-\psi}$.

Consider a bivariate random vector (X, Y), with marginal c.d.f.s for X and Y given by $F(x) = 1 - (1 + (x/\lambda)^\gamma)^{-\psi}$ and $F(y) = 1 - (1 + (y/\lambda)^\gamma)^{-\psi}$, respectively. The joint c.d.f. $F(x, y)$ is specified by the heavy right tail (HRT) copula given by

$$C(u, v) = u + v - 1 + \left[(1-u)^{-1/a} + (1-v)^{-1/a} - 1\right]^{-a}, \tag{5}$$

where $0 \le u \le 1$, $0 \le v \le 1$, and $a > 0$ (Frees and Valdez, 1998).

We set the copula parameter to be the same as the second shape parameter of the Burr distribution, that is, $a = \psi$. Replace u and v with $F(x)$ and $F(y)$, respectively, in (5). Then, the joint c.d.f. of the random vector (X, Y) is given by

$$\begin{aligned} F(x, y) &= F(x) + F(y) - 1 + \left[(1 - F(x))^{-1/\psi} + (1 - F(y))^{-1/\psi} - 1\right]^{-\psi} \\ &= 1 - \left(1 + \left(\frac{x}{\lambda}\right)^\gamma\right)^{-\psi} - \left(1 + \left(\frac{y}{\lambda}\right)^\gamma\right)^{-\psi} + \left[1 + \left(\frac{x}{\lambda}\right)^\gamma + \left(\frac{y}{\lambda}\right)^\gamma\right]^{-\psi}. \end{aligned}$$

The conditional c.d.f. of Y given $X = x$ is $F(y \,|\, x) = \partial C(F(x), F(y))/\partial F(x)$. Note that $\partial C(u, v)/\partial u = 1 - \left[(1-u)^{-1/\psi}\right]^{\psi+1} \left[(1-u)^{-1/\psi} + (1-v)^{-1/\psi} - 1\right]^{-(\psi+1)}$. It follows that

$$\begin{aligned} F(y \,|\, x) &= 1 - \left[1 + \left(\frac{x}{\lambda}\right)^\gamma\right]^{\psi+1} \left[1 + \left(\frac{x}{\lambda}\right)^\gamma + 1 + \left(\frac{y}{\lambda}\right)^\gamma - 1\right]^{-(\psi+1)} \\ &= 1 - \left[1 + \frac{\left(\frac{y}{\lambda}\right)^\gamma}{1 + \left(\frac{x}{\lambda}\right)^\gamma}\right]^{-(\psi+1)} = 1 - \left[1 + \frac{y^\gamma}{\lambda^\gamma + x^\gamma}\right]^{-(\psi+1)} \\ &= 1 - \left[1 + \left(\frac{y}{\tilde{\lambda}(x)}\right)^\gamma\right]^{-(\psi+1)}, \end{aligned} \tag{6}$$

where $\tilde{\lambda}(x) = (\lambda^\gamma + x^\gamma)^{1/\gamma}$. Therefore, the conditional distribution of Y given $X = x$ is a Burr distribution, $\mathrm{Burr}(\gamma, \tilde{\lambda}(x), \psi+1)$. Since the HRT copula is symmetric in its arguments, the conditional distribution of X given $Y = y$ is also a Burr distribution.

We note that the bivariate Burr distribution defined through the HRT copula and Burr marginals was considered in Venter (2002). However, the expressions for the conditional c.d.f.s reported in Venter (2002) include an error. Equation (6) provides the corrected expression for the conditional c.d.f. of Y given X.

C Implementation details

We outline the posterior simulation steps for the Burr MTDPP and the Lomax MTDCPP models illustrated in the data examples. Given an observed point pattern $\{t_i\}_{i=1}^n$ over the time window $(0, T)$, we let $x_1 = t_1$ and $x_i = t_i - t_{i-1}$ for $i = 2, \ldots, n$. For notational convenience, we take $x_{n+1} = T - t_n$. Our posterior inference is based on a likelihood, conditional on $(x_1, \ldots, x_L)$. Posterior samples of model parameters and latent variables are obtained by iteratively sampling from their posterior full conditional distributions.

Burr MTDPP We associate each x_i with a latent variable ℓ_i such that $P(\ell_i = l) = \sum_{l=1}^L w_l \delta_l(\ell_i)$. With customary priors for $(\gamma, \lambda, \kappa)$, we obtain the joint distribution

$$\mathrm{Ga}(\lambda \mid u_\lambda, v_\lambda) \times \mathrm{Ga}(\gamma \mid u_\gamma, v_\gamma) \times \mathrm{Ga}(\kappa \mid u_\kappa, v_\kappa) \mathbb{1}(\kappa > 1) \times \mathrm{Dir}(\boldsymbol{w} \mid \alpha_0 a_1, \ldots, \alpha_0 a_L)$$
$$\times \left\{ \prod_{i=L+1}^{n} \mathrm{Burr}\big(x_i \mid \gamma, \tilde{\lambda}(x_{i-\ell_i}), \kappa\big) \sum_{l=1}^{L} w_l \delta_l(\ell_i) \right\} \left\{ S\big(x_{n+1} \mid \gamma, \tilde{\lambda}(x_{n+1-\ell_{n+1}}), \kappa\big) \sum_{l=1}^{L} w_l \delta_l(\ell_{n+1}) \right\},$$

where $\tilde{\lambda}(v) = (\lambda^\gamma + v^\gamma)^{1/\gamma}$, and $S\big(x_{n+1} \mid \gamma, \tilde{\lambda}(x_{n+1-\ell_i}), \kappa\big)$ is the survival function associated the Burr distribution. In particular, $S\big(x_{n+1} \mid \gamma, \tilde{\lambda}(x_{n+1-\ell_i}), \kappa\big) = (1 + \{x_{n+1}/\tilde{\lambda}(x_{n+1-\ell_i})\}^\gamma)^{-\kappa}$.

Let $M_l = |\{i : \ell_i = l\}|$, for $l = 1, \ldots, L$. The posterior full conditional distribution of $\boldsymbol{w}$ is $\mathrm{Dir}(\boldsymbol{w} \mid \alpha_0 \alpha_1 + M_1, \ldots, \alpha_0 \alpha_L + M_L)^\top$. The posterior full conditional distribution of ℓ_i is a discrete distribution on $\{1, \ldots, L\}$, with probabilities proportional to $w_l \mathrm{Burr}\big(x_i \mid \gamma, \tilde{\lambda}(x_{i-l}), \kappa\big)$, for $i = L+1, \ldots, n$, and with probabilities proportional to $w_l S\big(x_{n+1} \mid \gamma, \tilde{\lambda}(x_{n+1-l}), \kappa\big)$ for $i = n+1$. Let $\boldsymbol{x} = (x_1, \ldots, x_{n+1})^\top$ and $\boldsymbol{\theta} = \{\lambda, \gamma, \kappa, \{\ell_i\}_{i=L+1}^{n+1}\}$. Take $p(\boldsymbol{x}, \boldsymbol{\theta}) = \left\{\prod_{i=L+1}^n \mathrm{Burr}(x_i \mid \gamma, \tilde{\lambda}(x_{i-\ell_i}), \kappa)\right\} S(x_{n+1} \mid \gamma, \tilde{\lambda}(x_{n+1-\ell_{n+1}}), \kappa)$. To update λ and γ, we use random walk Metropolis steps with target densities, respectively, $\mathrm{Ga}(\lambda \mid u_\lambda, v_\lambda)\, p(\boldsymbol{x}, \boldsymbol{\theta})$ and $\mathrm{Ga}(\gamma \mid u_\gamma, v_\gamma)\, p(\boldsymbol{x}, \boldsymbol{\theta})$. The posterior full conditional distribution of κ is a truncated gamma distribution, $\mathrm{Ga}(\kappa \mid u_\kappa + n - L, v_\kappa + \sum_{i=L+1}^{n+1} \log(1 + \{x_i/\tilde{\lambda}(x_{i-\ell_i})\}^\gamma)) \mathbb{1}(\kappa > 1)$.

Extended scaled-Lomax MTDPP Let $x_i = \mu(t_i)z_i$, with $\log\mu(t_i) = \sum_{j=1}^{J}\{\beta_{1j}\sin(j\omega t_i)+\beta_{2j}\cos(j\omega t_i)\}$. The conditional duration density is

$$f^*(x_i) = \mu(t_i)^{-1}\sum_{l=1}^{L} w_l P\big(\mu(t_i)^{-1}x_i \,|\, \alpha\phi + \mu(t_{i-l})^{-1}x_{i-l}, \alpha\big),$$

and the conditional survival function is $S^*(x_i) = \sum_{l=1}^{L} w_l S(\mu(t_i)^{-1}x_i \,|\, \alpha\phi+\mu(t_{i-l})^{-1}x_{i-l}, \alpha)$, where S is the survival function associated with the scaled-Lomax distribution. In particular, $S(z_i \,|\, \alpha\phi + z_{i-l}, \alpha) = (1 + \{\alpha\phi + z_{i-l}\}^{-1}z_i)^{-\alpha}$.

Denote $\boldsymbol{\beta} = (\beta_{11}, \ldots, \beta_{1J}, \beta_{21}, \ldots, \beta_{2J})^\top$, and let $\tilde{\beta}_k$ be the kth component of $\boldsymbol{\beta}$, for $k = 1, \ldots, 2J$. We introduce a collection of configuration variables $\{\ell_i\}_{i=L+1}^{n+1}$ such that $P(\ell_i = l) = \sum_{l=1}^{L} w_l\delta_l(\ell_i)$. With customary priors for $(\boldsymbol{\beta}, \alpha, \phi)$, we obtain the joint distribution

$$\prod_{k=1}^{2J} N(\tilde{\beta}_k \,|\, \mu_{\tilde{\beta}_k}, \sigma^2_{\tilde{\beta}_k}) \times \mathrm{Ga}(\alpha \,|\, u_\alpha, v_\alpha)\mathbb{1}(\alpha > 1) \times \mathrm{Ga}(\phi \,|\, u_\phi, v_\phi) \times \mathrm{Dir}(\boldsymbol{w} \,|\, \alpha_0 a_1, \ldots, \alpha_0 a_L)$$
$$\times \left\{ \prod_{i=L+1}^{n} \mu(t_i)^{-1} P\left(\mu(t_i)^{-1}x_i \,|\, \alpha\phi + \mu(t_{i-\ell_i})^{-1}x_{i-\ell_i}, \alpha\right) \sum_{l=1}^{L} w_l\delta_l(\ell_i) \right\}$$
$$\times \left\{ S\big(\mu(t_{n+1})^{-1}x_{n+1} \,|\, \alpha\phi + \mu(t_{n+1-\ell_{n+1}})^{-1}x_{n+1-\ell_{n+1}}, \alpha\big) \sum_{l=1}^{L} w_l\delta_l(\ell_{n+1}) \right\},$$

where, by abuse of notation, we take $t_{n+1} = T$.

The posterior full conditional distributions for the weights $\boldsymbol{w}$ and configuration variables $\{\ell_i\}_{i=L+1}^{n+1}$ are similar to those for the Burr MTDPP. We focus on the updates for $(\boldsymbol{\beta}, \alpha, \phi)$. Let $\boldsymbol{t} = (t_1, \ldots, t_{n+1})^\top$, $\boldsymbol{x} = (x_1, \ldots, x_{n+1})^\top$, and $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \alpha, \phi, \{\ell_i\}_{i=L+1}^{n+1}\}$. We take

$$p(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{\theta}) = \left\{ \prod_{i=L+1}^{n} P\left(\mu(t_i)^{-1}x_i \,|\, \alpha\phi + \mu(t_{i-\ell_i})^{-1}x_{i-\ell_i}, \alpha\right) \right\}$$
$$\times S\big(\mu(t_{n+1})^{-1}x_{n+1} \,|\, \alpha\phi + \mu(t_{n+1-\ell_{n+1}})^{-1}x_{n+1-\ell_{n+1}}, \alpha\big).$$

We use random walk Metropolis steps to update α and ϕ with target densities, respectively, $\mathrm{Ga}(\alpha \,|\, u_\alpha, v_\alpha)\mathbb{1}(\alpha > 1)\, p(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{\theta})$ and $\mathrm{Ga}(\phi \,|\, u_\phi, v_\phi)\, p(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{\theta})$. Similarly, we update $\tilde{\beta}_k$ using a random walk Metropolis step with target density $N(\tilde{\beta}_k \,|\, \mu_{\tilde{\beta}_k}, \sigma^2_{\tilde{\beta}_k})\, p(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{\theta}) \prod_{i=L+1}^{n+1} \mu(t_i)^{-1}$, for $k = 1, \ldots, 2J$.

Lomax MTDCPP The Lomax MTDCPP conditional duration density, for $i > L$, can be written as $f_C^*(x_i) = \sum_{l=0}^{L} \pi_l f_l^c(x_i \,|\, \mu, \phi, \alpha)$, where $f_0^c \equiv \mu\exp(-\mu x_i)$, $f_l^c \equiv P(x_i \,|\, \phi+x_{i-l}, \alpha)$, and $\pi_l = (1-\pi_0)w_l,\ l = 1,\ldots,L$. Let S_0^c and S_l^c be the survival functions associated with f_0^c and f_l^c, respectively. With customary priors for (μ,ϕ,α), the joint distribution is

$$\mathrm{Ga}(\mu\,|\,u_\mu,v_\mu)\times\mathrm{Ga}(\phi\,|\,u_\phi,v_\phi)\times\mathrm{Ga}(\alpha\,|\,u_\alpha,v_\alpha)\mathbb{1}(\alpha>1)\times\mathrm{Dir}(\boldsymbol{w}\,|\,\alpha_0a_1,\ldots,\alpha_0a_L)$$
$$\times\,\mathrm{Beta}(\pi_0\,|\,u_0,v_0)\times\left\{\prod_{i=L+1}^{n} f_{\ell_i}^c(x_i\,|\,\mu,\phi,\alpha)\sum_{l=0}^{L}\pi_l\delta_l(\ell_i)\right\}\left\{S_{\ell_{n+1}}^c(x_{n+1}\,|\,\mu,\phi,\alpha)\sum_{l=0}^{L}\pi_l\delta_l(\ell_{n+1})\right\}.$$

The posterior full conditional distributions for $\boldsymbol{w}$, π_0, and $\{\ell_i\}_{i=L+1}^{n+1}$ can be found in the main paper. We focus on the posterior updates for (μ,ϕ,α). Let $M_l = |\{i : \ell_i = l\}|$, for $l = 0,\ldots,L$. A gamma prior for μ yields conjugate full conditional distribution $\mathrm{Ga}(\mu\,|\,u_\mu + M_0 - \delta_0(\ell_{n+1}), v_\mu + \sum_{i:\ell_i=0} x_i)$. The full conditional distribution of α is truncated gamma distribution $\mathrm{Ga}(\alpha\,|\,u_\alpha + \sum_{l=1}^{L} M_l - 1 + \delta_0(\ell_{n+1}), v_\alpha + \sum_{i:\ell_i\neq 0}\log(1 + x_i/(\phi + x_{i-\ell_i}))\mathbb{1}(\alpha>1)$. Let $B_0^n = \{i : (1\le i\le n)\wedge(\ell_i\neq 0)\}$, and $p(\{x_i\}_{i=1}^{n+1},\mu,\phi,\alpha,\{\ell_i\}_{i=L+1}^{n+1}) = \left\{\prod_{i\in B_0^n} f_{\ell_i}^c(x_i\,|\,\mu,\phi,\alpha)\right\}\left\{S_{\ell_{n+1}}^c(x_{n+1}\,|\,\mu,\phi,\alpha)\right\}^{1-\delta_0(\ell_{n+1})}$. We update ϕ using a random walk Metropolis step with target density $\mathrm{Ga}(\phi\,|\,u_\phi,v_\phi)\,p(\{x_i\}_{i=1}^{n+1},\mu,\phi,\alpha,\{\ell_i\}_{i=L+1}^{n+1})$.

D Additional simulation experiment

The goal of this experiment is to examine the ability of the MTDCPP to recover various clustering behaviors attributed to two different factors. To this end, we generate data from a Lomax MTDCPP, that is, with f_I corresponding to an exponential distribution with rate parameter μ and $f^*(t - t_{N(t)})$ a stationary Lomax MTDPP.

In the experiment, we consider four scenarios, with π_0 taking one of the following values, (0.2, 0.5, 0.8, 1). The first three values indicate the proportion of the duration process affected by a factor through f_I. When $\pi_0 = 1$, the data are equivalently generated by a Poisson process with rate μ. For all scenarios, we take $\mu = 0.2$, $\alpha = 5$, $\phi = 0.1$ and decaying weights $\boldsymbol{w} = (0.35, 0.25, 0.2, 0.1, 0.1)^\top$.

We applied the Lomax MTDCPP model with $L = 5$ to the synthetic data. We specified a beta prior $\mathrm{Beta}(\pi_0\,|\,1,1)$ for the probability π_0 and a gamma prior $\mathrm{Ga}(\mu\,|\,1,1)$ for the rate parameter μ of the exponential distribution f_I. For the stationary Lomax MTDPP, the shape and scale parameters received priors $\mathrm{Ga}(\alpha\,|\,6,1)\mathbb{1}(\alpha>1)$ and $\mathrm{Ga}(\phi\,|\,1,1)$. In particular, we chose prior for α with the expectation that the first four moments exist with respect to the component and marginal Lomax distributions. The vector $\boldsymbol{w}$ was assigned $\mathrm{CDP}(\boldsymbol{w}\,|\,5,1,3)$, which elicits a decreasing pattern in the weights.

Table 1: Additional simulation experiment. Posterior means and 95% credible interval estimates of the MTDCPP model parameters under different scenarios.

	$\pi_0 = 0.2$	$\pi_0 = 0.5$	$\pi_0 = 0.8$	$\pi_0 = 1$
π_0	0.22 (0.20, 0.25)	0.52 (0.48, 0.56)	0.81 (0.76, 0.86)	0.99 (0.96, 1.00)
μ	0.22 (0.19, 0.24)	0.19 (0.17, 0.20)	0.20 (0.19, 0.21)	0.19 (0.18, 0.20)
ϕ	0.12 (0.09, 0.15)	0.13 (0.09, 0.19)	0.12 (0.02, 0.33)	1.33 (0.04, 4.70)
α	5.46 (4.63, 6.43)	6.37 (4.91, 8.15)	4.78 (2.99, 8.21)	5.39 (2.19, 10.82)

We focus on the inference on the two-component mixture probability π_0 and the component density parameters (μ, ϕ, α). The posterior means and 95% credible interval estimates of the parameters are presented in Table 1. The posterior estimates of the parameter π_0 suggest that the model was able to recover the proportion of the point process driven by f_I, even in the extreme case when $\pi_0 = 1$. For other parameters, the model produced estimates close to the true values for all scenarios.

E Model checking results

We provide model checking results for the data examples. Define $U_i^* = 1 - \exp\{-(\Lambda^*(t_i) - \Lambda^*(t_{i-1}))\} = F^*(t_i - t_{i-1})$ as residuals of the point process model. If the model is correctly specified, the residuals will be independently and identically distributed as a standard uniform distribution. Figure 1 consists of quantile-quantile plots of the estimated U_i^* for the simulation experiment and the first real data example of the paper, and the additional experiment in Section D. Figures 2-5 contain quantile-quantile plots of the estimated U_i^* for the second real data example of the paper. The graphical model assessment results indicate good model fit for all data examples.

References

Arnold, B. C., Castillo, E., and Sarabia, J. M. (1999), *Conditional Specification of Statistical Models*, New York: Springer.

Frees, E. W. and Valdez, E. A. (1998), "Understanding relationships using copulas," *North American actuarial journal*, 2, 1–25.

Korchevsky, V. and Petrov, V. (2010), "On the strong law of large numbers for sequences of dependent random variables," *Vestnik St. Petersburg University: Mathematics*, 43, 143–147.

Tadikamalla, P. R. (1980), "A look at the Burr and related distributions," *International Statistical Review/Revue Internationale de Statistique*, 337–344.

Venter, G. G. (2002), "Tails of copulas," in *Proceedings of the Casualty Actuarial Society*, volume 89.

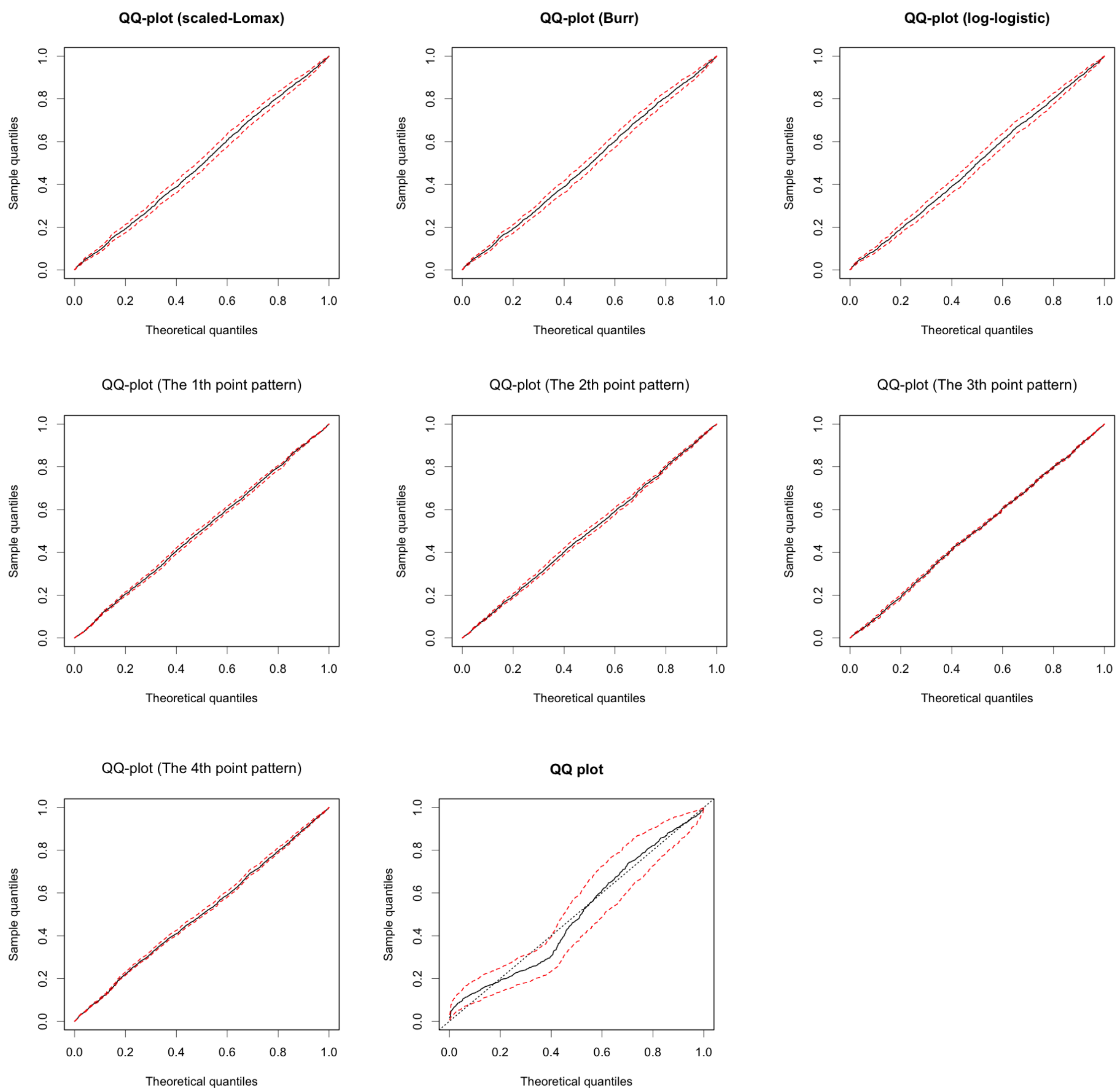


Figure 1: Model checking results. The first row corresponds to the simulation experiment in the main paper. The second row and the first panel of the third row correspond to the additional simulation experiment. The second panel of the third row corresponds to the IVT data example. Black solid lines are posterior means and red dotted lines are 95% credible interval estimates.

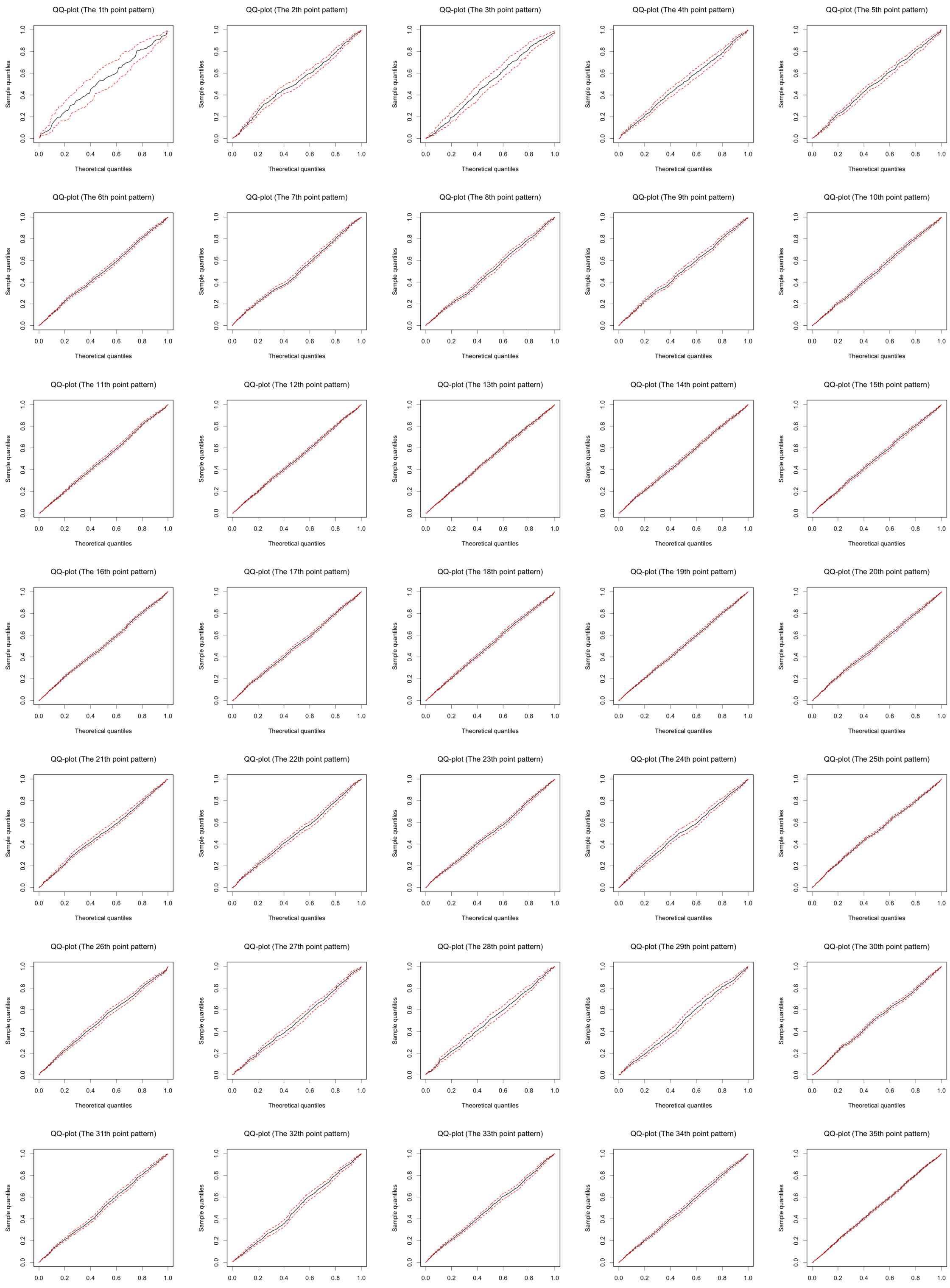


Figure 2: Model checking results. Forex data example point patterns 1 - 35. Black solid lines are posterior means and red dotted lines are 95% credible interval estimates.

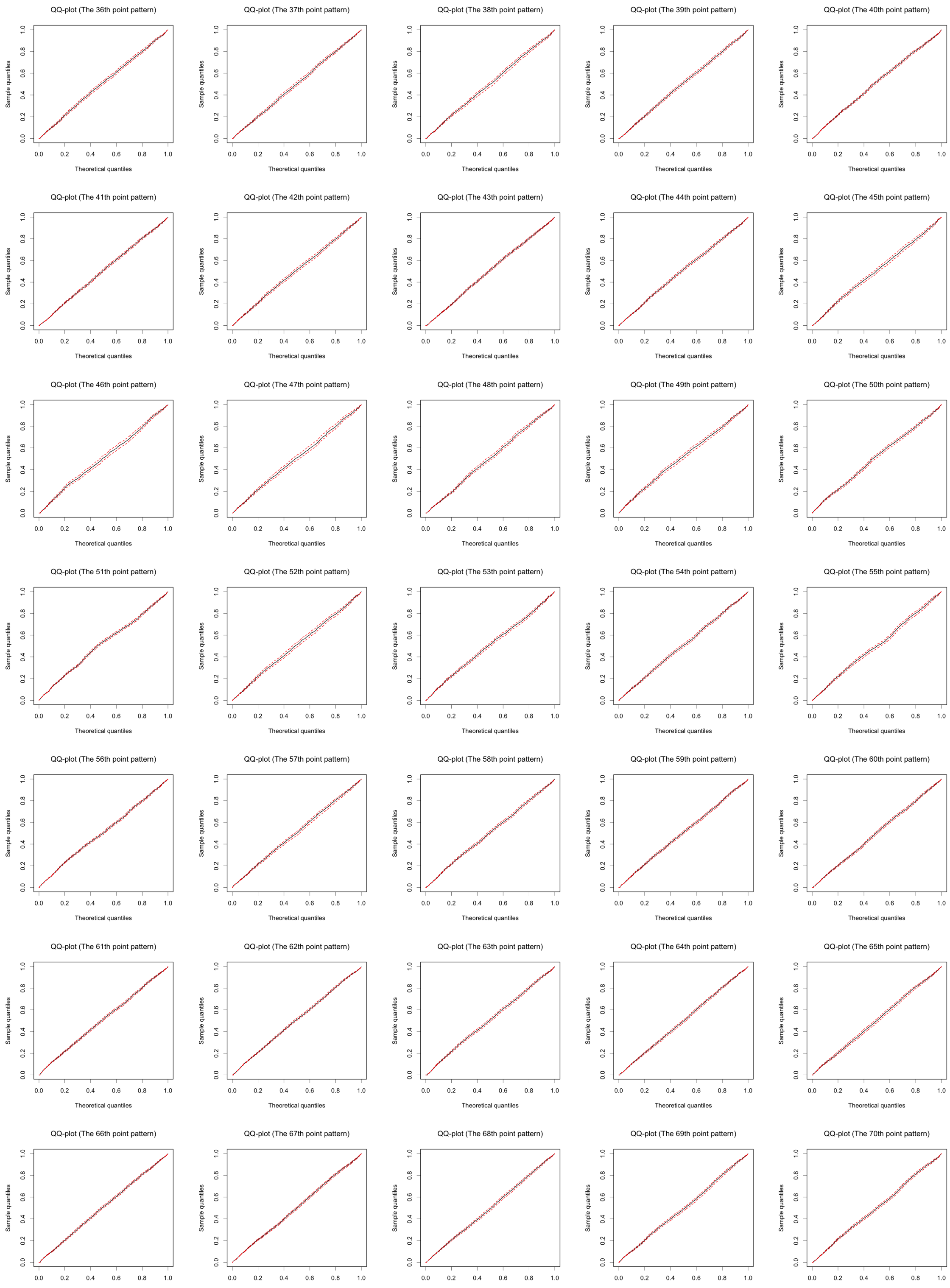


Figure 3: Model checking results. Forex data example point patterns 36 - 70. Black solid lines are posterior means and red dotted lines are 95% credible interval estimates.

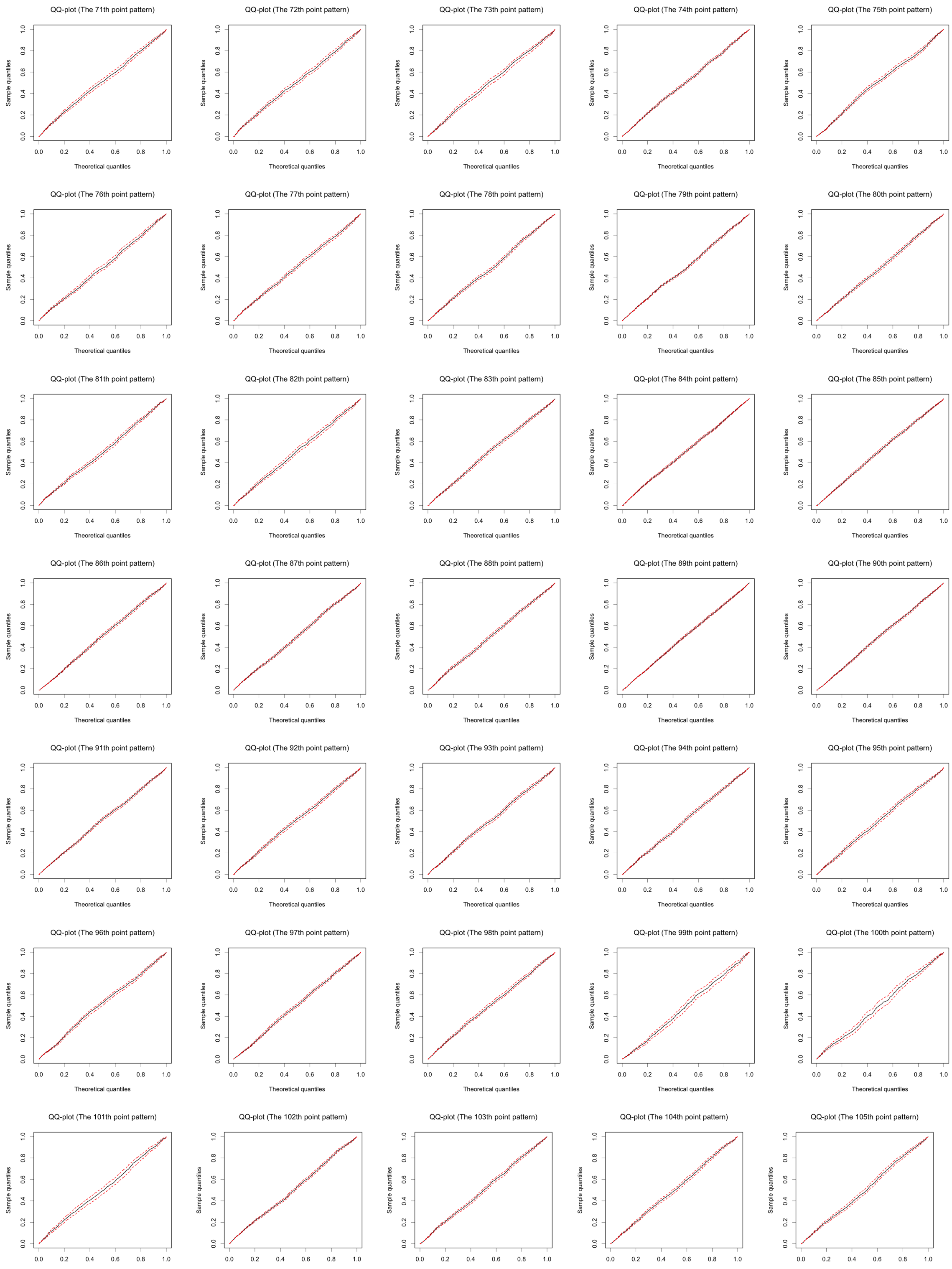


Figure 4: Model checking results. Forex data example point patterns 71 - 105. Black solid lines are posterior means and red dotted lines are 95% credible interval estimates.

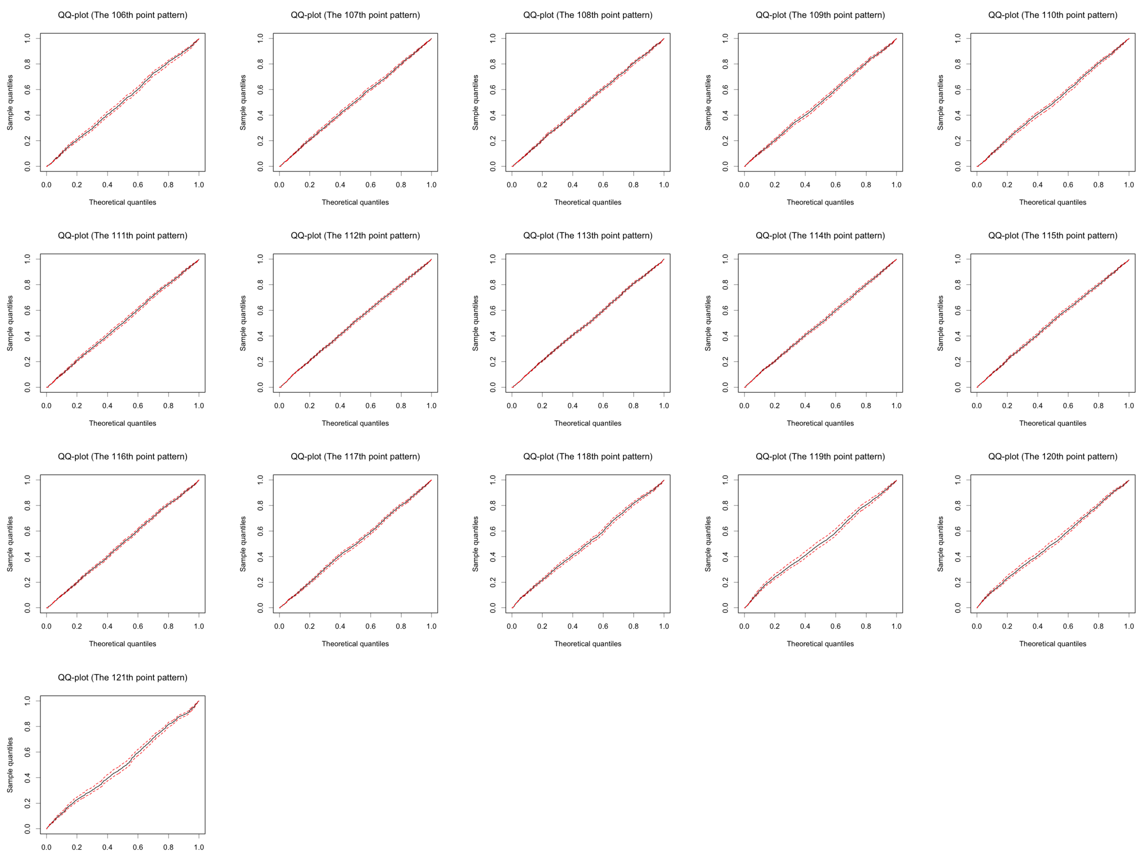


Figure 5: Model checking results. Forex data example point patterns 106 - 121. Black solid lines are posterior means and red dotted lines are 95% credible interval estimates.