DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Inference and forecasting for continuous-time integer-valued trawl processes and their use in financial economics

Mikkel Bennedsen, Asger Lunde, Neil Shephard and Almut E. D. Veraart

CREATES Research Paper 2021-12

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 9352 1936

Inference and forecasting for continuous-time integer-valued trawl processes and their use in financial economics

Mikkel Bennedsen[*], Asger Lunde[†], Neil Shephard[‡], Almut E. D. Veraart[§]

July 9, 2021

Abstract

This paper develops likelihood-based methods for estimation, inference, model selection, and forecasting of continuous-time integer-valued trawl processes. The full likelihood of integer-valued trawl processes is, in general, highly intractable, motivating the use of composite likelihood methods, where we consider the pairwise likelihood in lieu of the full likelihood. Maximizing the pairwise likelihood of the data yields an estimator of the parameter vector of the model, and we prove consistency and asymptotic normality of this estimator. The same methods allow us to develop probabilistic forecasting methods, which can be used to construct the predictive distribution of integer-valued time series. In a simulation study, we document good finite sample performance of the likelihood-based estimator and the associated model selection procedure. Lastly, the methods are illustrated in an application to modelling and forecasting financial bid-ask spread data, where we find that it is beneficial to carefully model both the marginal distribution and the autocorrelation structure of the data. We argue that integer-valued trawl processes are especially well-suited in such situations.

JEL codes: C01; C13; C22; C51; C53; G17.

Keywords: Integer valued trawl process; Lévy basis; composite likelihood; pairwise likelihood; estimation; model selection; forecasting.

[*]Department of Economics and Business Economics and CREATES, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark. E-mail: mbennedsen@econ.au.dk.

[†]Copenhagen Economics, Langebrogade 1B, 1411 Copenhagen K, Denmark, and CREATES, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark. E-mail: alu@CopenhagenEconomics.com.

[‡]Department of Economics and Department of Statistics, Harvard University, One Oxford Street, Cambridge, MA 02138, USA. E-mail: shephard@fas.harvard.edu.

[§]Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK and CREATES, Aarhus University. E-mail: a.veraart@imperial.ac.uk.

1 Introduction

Count-valued time series are encountered in many fields, such as epidemiology (Wakefield, 2007; Schmidt and Pereira, 2011; Pedeli et al., 2015), environmental studies (Livsey et al., 2018; Noven et al., 2018), and finance (Fokianos et al., 2009, 2020; Barndorff-Nielsen et al., 2012; Pedeli and Karlis, 2013; Shephard and Yang, 2017; Veraart, 2019). Developing methods for estimation and inference for count-valued models is thus an important endeavour, and many approaches have been suggested in the literature. For instance, Al-Osh and Alzaid (1987) proposes several methods for estimating so-called first order integer-valued autoregressive (INAR(1)) models, while Pedeli et al. (2015) suggest saddlepoint approximation methods for estimating INAR(p) models, and Fokianos (2016) considers a GLM perspective on estimation. Recently, spectral estimation and indirect inference have been pursued in Doukhan et al. (2020) and Davis et al. (2020), respectively, while Doukhan et al. (2021) estimates Markov switching integer-valued ARCH models using maximum likelihood methods. For extensive treatments of count-valued time series, we refer to Davis et al. (2016) and Karlis (2016). A recent review is Davis et al. (2021).

In this paper, we develop likelihood-based methods for estimation, inference, model selection, and forecasting of continuous-time integer-valued trawl (IVT) processes. IVT processes, introduced in Barndorff-Nielsen et al. (2014), are a flexible class of integer-valued, serially correlated, stationary, and infinitely divisible continuous-time stochastic processes. In general, however, IVT processes are not Markovian, which implies that the structure of the full likelihood of an IVT process is highly intractable (Shephard and Yang, 2016). This is the impetus of the present paper, where we propose to use composite likelihood (CL, Lindsay, 1988) methods for estimation and inference. Specifically, we propose to estimate the parameters of an IVT model by maximizing a pseudo/quasi likelihood function constructed using only pairs of data points, the so-called pairwise likelihood. CL methods in general, and the pairwise likelihood approach in particular, have been succesfully used in many applications, such as statistical genetics (Larribe and Fearnhead, 2011), geostatistics (Hjort and Omre, 1994), and finance (Engle et al., 2020). Although the theory behind CL estimation is quite well understood in the case of iid observations (e.g. Cox and Reid, 2004; Varin and Vidoni, 2005; Varin, 2008), the time series case, which is what we consider here, requires separate treatment (Varin et al., 2011, p. 11).

A central feature of IVT processes is that they allow for specifying the correlation structure of the model separately from the marginal distribution of the model, making them flexible and well-suited for modelling count- or integer-valued data. In particular, the marginal distribution of an IVT process can be any integer-valued infinitely divisible distribution, while the correlation structure can be specified independently using a so-called trawl function. This setup allows for both short- and long-memory of the IVT process. So far, IVT processes have been applied to financial data (Barndorff-Nielsen et al., 2014; Shephard and Yang, 2017; Veraart, 2019) and to the modelling of extreme events in environmental time series (Noven et al., 2018). IVT processes are,

under weak conditions, stationary and ergodic, which motivated Barndorff-Nielsen et al. (2014) to suggest a method of moments-based estimator for the parameters of the IVT model. This method of moments estimator has been used in most applied work using IVT processes (e.g. Barndorff-Nielsen et al., 2014; Shephard and Yang, 2017; Veraart, 2019). Exceptions are Shephard and Yang (2016) and Noven et al. (2018). In Noven et al. (2018), a pairwise likelihood was used for a hierarchical model involving a latent (Gamma-distributed) trawl process and the corresponding asymptotic theory was derived in Courgeau and Veraart (2020). However, the asymptotic theory for inference for integer-valued trawl processes which are observed directly, is not covered by these earlier papers. In Shephard and Yang (2016), the authors derive a prediction decomposition of the likelihood function of a particularly simple IVT process, the so-called Poisson-Exponential IVT process, allowing them to conduct likelihood-based estimation and inference. Although the likelihood estimation method developed in Shephard and Yang (2016) theoretically applies to more general IVT processes, the computational burden quickly becomes overwhelming in these scenarios, making estimation by classical maximum likelihood methods infeasible in practice.

The contributions of this paper can be summarized as follows. First, we prove consistency and asymptotic normality of the maximum composite likelihood (MCL) estimator of the parameter vector of an IVT model. For the purpose of conducting feasible inference and model selection, we propose two alternative estimators of the asymptotic variance of the MCL estimator: a kernel-based estimator, inspired by the heteroskedastic and autocorrelation consistent (HAC) estimator of Newey and West (1987), and a simulation-based estimator. Second, we use the same principle of considering the pairwise likelihood in lieu of the full likelihood, to derive the predictive distribution of an IVT model, conditional on the current value of the process; this allows us to use the IVT framework for forecasting integer-valued data. In a simulation study, we compare the MCL estimator to the standard method of moments-based estimator suggested in Barndorff-Nielsen et al. (2014) and find that the MCL estimator provides substantial improvements in most cases. Indeed, in a realistic simulation setup, we find that the MCL estimator can improve on the method-of-moments-based estimator by more than 50%, in terms of finite sample root median squared error.

We apply the methods developed in the paper to a time series of the bid-ask spread of a financial asset. The time series behavior of the bid-ask spread has been extensively studied in the literature on the theory of the microstructure of financial markets (e.g. Huang and Stoll, 1997; Bollen et al., 2004). The model selection procedure developed in the paper indicates that a model with Negative Binomial marginal distribution and slowly decaying autocorrelations most adequately describe the data. These findings are in line with those of Groß-KlußMann and Hautsch (2013), who also found strong persistence in bid-ask spread time series. Then, in a pseudo out-of-sample forecast exercise, we find that it is important to carefully model both the marginal distribution and the autocorrelation structure to get accurate forecasts of the future bid-ask spread. These findings illustrate the strength of modelling using a framework where the choice of marginal distribution

can be made independently of the choice of autocorrelation structure.

The rest of the paper is structured as follows. Section 2 outlines the mathematical setup of IVT processes, while Section 3 contains details on the estimation and model selection procedures. Section 4 presents the theory behind our proposed forecasting approach. Section 5 contains an extensive Monte Carlo simulation study, investigating the finite sample properties of the estimation and model selection procedures. Section 6 illustrates the use of the methods of this paper in an empirical application to financial bid-ask spread data. Section 7 concludes. Practical details on implementation of the asymptotic theory, and additional derivations are given in an appendix. A Supplementary Material file, see Bennedsen et al. (2021), contains the proofs of the theoretical results, further simulation results and extensive details on various calculations used in the implementation of the methods. A software package for implementation of simulation, estimation, inference, model selection, and forecasting of IVT processes is freely available in the MATLAB programming language.[1]

2 Integer-valued trawl processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space, satisfying the usual assumptions and supporting a Poisson random measure N, defined on $\mathbb{Z} \times [0,1] \times \mathbb{R}$, with mean (intensity) measure $\eta \otimes Leb \otimes Leb$. Here Leb denotes the Lebesgue measure and η is a Lévy measure. That is,

$$\mathbb{E}[N(dy, dx, ds)] = \eta(dy)dxds. \tag{2.1}$$

We further assume that $\|\eta\| := \int_{-\infty}^{\infty} y\eta(dy) < \infty$. A *Lévy basis* L can be constructed on $[0,1] \times \mathbb{R}$ by defining

$$L(dx, ds) := \int_{-\infty}^{\infty} yN(dy, dx, ds), \quad (x, s) \in [0,1] \times \mathbb{R}.$$

Intuitively, for a Borel set $B \in \mathcal{B}([0,1] \times \mathbb{R})$, $L(B)$ sums the number of events in B, weighted by the "size" of an event, y; the events are distributed uniformly over "height" ($x \in [0,1]$) and "time" ($s \in \mathbb{R}$), where the "size" of the events are distributed according to the measure $\eta(y)$. Since we are only interested in integer-valued Lévy bases, η is concentrated on the integers ($y \in \mathbb{Z}$).

The Lévy basis L is an infinitely divisible random measure with cumulant (log-characteristic) function

$$C_{L(dx,ds)}(\theta) := \log \mathbb{E}[\exp(i\theta L(dx, ds))] = \int_{-\infty}^{\infty} \left(e^{i\theta y} - 1\right) \eta(dy)dxds, \quad (x, s) \in [0,1] \times \mathbb{R}. \tag{2.2}$$

An important random variable associated with the Lévy basis L, is the so-called *Lévy seed*, L'.

[1] The software package can be found at https://github.com/mbennedsen/Likelihood-based-IVT.

Definition 1 (Lévy seed)**.** *Let L be an integer-valued, homogeneous Lévy basis with cumulant function given by (2.2). Then, the random variable L' satisfying $\mathbb{E}[\exp(i\theta L')] = \exp(C_{L'}(\theta))$, with $C_{L'}(\theta) = \int \left(e^{i\theta y} - 1\right) \eta(dy)$, is called the Lévy seed associated with the Lévy basis L.*

Remark 2.1. Because the distribution of a Lévy process is entirely determined by its distribution at a particular time point, we can specify a Lévy process L'_t from a Lévy seed L', by requiring that $L'_1 \sim L'$.

Using the Lévy seed, we can rewrite the cumulant function of the Lévy basis (2.2) as $C_{L(dx,ds)}(\theta) = C_{L'}(\theta)dxds$, or, for a Borel set $B \in \mathcal{B}([0,1] \times \mathbb{R})$,

$$C_{L(B)}(\theta) = C_{L'}(\theta)Leb(B). \tag{2.3}$$

From (2.3) we have that $\kappa_j(L(B)) = \kappa_j(L')Leb(B)$, $j \geq 0$, where $\kappa_j(Z)$ denotes the j'th cumulant of the random variable Z, when it exists.[2] In particular

$$\mathbb{E}[L(B)] = \mathbb{E}[L']Leb(B) \quad \text{and} \quad Var(L(B)) = Var(L')Leb(B). \tag{2.4}$$

The relationship (2.3) implies that the distribution of the random variable $L(B)$ is entirely specified by the Lévy seed L' and the Lebesgue measure of the set B. In Section 2.1 below, we illustrate how this can be used to construct trawl processes with a given marginal distribution.

The Lévy basis L acts on sets in $\mathcal{B}([0,1] \times \mathbb{R})$. We restrict attention to *trawl sets* of the form

$$A_t = A + (0,t), \quad A = \{(x,s) : s \leq 0, 0 \leq x < d(s)\}, \quad t \geq 0, \tag{2.5}$$

where $d : \mathbb{R}_- \to [0,1]$ is a *trawl function*, which determines the shape of the trawl set A_t. Section 2.2 elaborates on the assumptions we make on d. Intuitively, A_t is obtained from the set A by "dragging" it along in time. Note in particular that $Leb(A_t) = Leb(A)$ for all t. Finally, define the integer-valued trawl (IVT) process $X = (X_t)_{t \geq 0}$ as the Lévy basis evaluated over the trawl set:

$$X_t := L(A_t), \quad t \geq 0. \tag{2.6}$$

The following sections illustrate how to use this setup as a flexible modelling tool. In particular, we demonstrate how the marginal distribution (Section 2.1) and the correlation structure (Section 2.2) can be specified independently. Some specific examples are provided.

2.1 Modelling the marginal distribution

Let $B \in \mathcal{B}([0,1] \times \mathbb{R})$ be a Borel set which, in what follows, will often be a subset of the trawl set A_t. We consider the distribution of the Lévy basis evaluated over the set B, i.e. of the random variable $L(B)$. From (2.3) we have that

$$C_{L(B)}(\theta) = Leb(B)C_{L'}(\theta) = C_{L'_{Leb(B)}}(\theta),$$

[2] Recall that the cumulants $\kappa_j(Z)$ of the random variable Z are defined implicitly through the power series expansion of the cumulant function of Z, i.e., $C_Z(\theta) = \log \mathbb{E}[\exp(i\theta Z)] = \sum_{j=1}^{\infty} \kappa_j(Z)(i\theta)^j/j!$.

where L'_t is a Lévy process with $L'_1 \sim L'$. This implies that the distribution of the random variable $L(B)$ can be seen as those of the Lévy process L'_t, induced by the Lévy seed L', at time $t = Leb(B)$, see Remark 2.1. In particular, since $Leb(A_t) = Leb(A)$ for all t, we have

$$C_{X_t}(\theta) = C_{L(A_t)}(\theta) = Leb(A)C_{L'}(\theta) = C_{L'_{Leb(A)}}(\theta),$$

where X_t is the IVT process defined in (2.6).

From this discussion it is clear that the marginal distribution of the IVT process X_t is entirely decided by the Lebesgue measure of the trawl set A and the Lévy seed L' of the underlying Lévy basis L. Indeed, by specifying a distribution for L', we can built IVT processes with the corresponding marginal distribution. The following two examples illustrates how to do this; additional details can be found in the Supplementary Material, where more examples are also given.

Example 2.1 (Poissonian Lévy seed)**.** Let $L' \sim \text{Poisson}(\nu)$, i.e. L' is distributed as a Poisson random variable with intensity $\nu > 0$. It follows from standard properties of the Poisson distribution that $X_t \sim \text{Poisson}(\nu Leb(A))$. In other words, for all $t \geq 0$,

$$P\left(X_t = x\right) = (\nu Leb(A))^x e^{-\nu Leb(A)}/x!, \quad x = 0, 1, 2, \ldots.$$

Example 2.2 (Negative Binomial Lévy seed)**.** Let $L' \sim \text{NB}(m, p)$, i.e. L' is distributed as a Negative Binomial random variable with parameters $m > 0$ and $p \in [0, 1]$. It follows from standard properties of the Negative Binomial distribution that $X_t \sim \text{NB}(mLeb(A), p)$. In other words, for all $t \geq 0$,

$$P(X_t = x) = \frac{\Gamma(Leb(A)m + x)}{x!\Gamma(Leb(A)m)}(1-p)^{Leb(A)m}p^x, \quad x = 0, 1, 2, \ldots,$$

where $\Gamma(z) = \int_0^\infty y^{z-1}e^{-y}dy$ for $z > 0$ is the Γ-function.

2.2 Modelling the correlation structure

Recall that the shape of the trawl set A_t is determined by the trawl function d, see Equation (2.5). We will restrict attention to a class of trawl functions with a particularly flexible structure, the so-called *superposition trawls* (Barndorff-Nielsen et al., 2014; Shephard and Yang, 2017). They are defined as

$$d(s) := \int_0^\infty e^{\lambda s}\pi(d\lambda), \quad s \leq 0,$$

where π is a probability measure on $\mathbb{R}_+$. This construction essentially randomizes the decay parameter λ in an otherwise exponential function. Given a trawl function d, it holds that

$$Leb(A_t) = Leb(A) = \int_{-\infty}^0 d(s)ds. \tag{2.7}$$

It is also useful to note that for $t \geq s$,

$$Leb(A_t \cap A_s) = Leb(A_{t-s} \cap A) = \int_{-\infty}^{-(t-s)} d(s)ds, \tag{2.8}$$

and

$$Leb(A_t \setminus A_s) = Leb(A_s \setminus A_t) = L(A) - Leb(A_t \cap A_s). \tag{2.9}$$

Thus, given a trawl function d, it is straightforward to calculate $Leb(A)$, $Leb(A_t \cap A_s)$, and $L(A_t \setminus A_s)$ for all $t \geq s$.

As we saw above, the Lévy seed, together with the Lebesgue measure of A, controls the marginal distribution of the IVT process. Analogously, the trawl function d controls the autocorrelation structure of the process. To see this, note that for $t, s \geq 0$ (Barndorff-Nielsen et al., 2014),

$$Cov(L(A_t), L(A_s)) = Var(L')Leb(A_t \cap A_s) = Var(L')Leb(A_{t-s} \cap A).$$

The IVT process with a superposition trawl function is stationary, so using (2.4) we get the correlation function for $h > 0$,

$$\rho(h) := Corr(L(A_{t+h}), L(A_t)) = \frac{Leb(A_h \cap A)}{Leb(A)} = \frac{\int_h^\infty d(-s)ds}{\int_0^\infty d(-s)ds}, \tag{2.10}$$

from which it is evident how the trawl function d directly determines the correlation structure of the trawl process. The following three examples illustrate this approach; additional details can be found in the Supplementary Material, where more examples are also given.

Example 2.3 (Exponential trawl function)**.** For the case where the measure π has an atom at $\lambda > 0$, i.e. $\pi(dx) = \delta_\lambda(dx)$, where $\delta_x(\cdot)$ is the Dirac delta function at $x \in \mathbb{R}_+$, we get $d(s) = e^{\lambda s}$ for $s \leq 0$. Consequently, the correlation function of the IVT process with exponential trawl function becomes

$$\rho(h) = Corr(X_{t+h}, X_t) = \frac{\int_h^\infty d(-s)ds}{\int_0^\infty d(-s)ds} = \exp(-\lambda h), \quad h \geq 0.$$

Example 2.4 (Inverse Gaussian trawl function)**.** Letting π be given by the inverse Gaussian distribution

$$\pi(dx) = \frac{(\gamma/\delta)^{1/2}}{2K_{1/2}(\delta\gamma)} x^{-1/2} \exp\left(-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right) dx,$$

where $K_\nu(\cdot)$ is the modified Bessel function of the third kind and $\gamma, \delta \geq 0$ with both not zero simultaneously. It can be shown that the resulting trawl function is given by

$$d(s) = \left(1 - \frac{2s}{\gamma^2}\right)^{-1/2} \exp\left(\delta\gamma\left(1 - \sqrt{1 - \frac{2s}{\gamma^2}}\right)\right), \quad s \leq 0,$$

and hence that the correlation function of the IVT process with inverse Gaussian trawl function becomes

$$\rho(h) = Corr(X_{t+h}, X_t) = \exp\left(-\delta\gamma(\sqrt{1+2h/\gamma^2}) - 1\right), \quad h \geq 0.$$

The details on these calculations can be found in the Supplementary Material.

Example 2.5 (Gamma trawl function)**.** Let π have the $\Gamma(1+H, \alpha)$ density,

$$\pi(dx) = \frac{1}{\Gamma(1+H)}\alpha^{1+H}\lambda^H e^{-\lambda\alpha}dx,$$

where $\alpha > 0$ and $H > 0$. We can show that

$$d(s) = \left(1 - \frac{s}{\alpha}\right)^{-(H+1)}, \quad s \leq 0,$$

which implies the correlation function

$$\rho(h) = Corr(X_{t+h}, X_t) = \frac{Leb(A_h \cap A)}{Leb(A)} = \left(1 + \frac{h}{\alpha}\right)^{-H}.$$

Note that in this case

$$\int_0^\infty \rho(h)dh = \begin{cases} \infty & \text{if} \quad H \in (0,1], \\ \frac{\alpha}{H-1} & \text{if} \quad H > 1, \end{cases}$$

from which we see that an IVT process with a Gamma trawl function enjoys the long memory property, in the sense of a non-integrable autocorrelation function, when $H \in (0, 1]$. The details on these calculations can be found in the Supplementary Material.

2.3 Modelling IVT processes

Using the above methods, we can build flexible continuous-time integer-valued processes with a marginal distribution determined by the underlying Lévy basis, and independently specified correlation structure determined by the trawl function. In our main examples given above, we considered a Lévy basis with Poisson (Example 2.1) or Negative Binomial (Example 2.2) marginals, and various trawl functions, namely the Exponential trawl function (Example 2.3), the IG trawl function (Example 2.4), and the Gamma trawl function (Example 2.5). Other specifications for the underlying Lévy basis and trawl function than those given here could of course be considered. In practice, these choices should be guided by the properties of the data being modelled.

The simplest IVT process we can construct in this way is the Poisson-Exponential IVT process, i.e., the the case where $L' \sim \text{Poisson}(\nu)$ and $d(s) = \exp(\lambda s)$, $s \leq 0$, see Examples 2.1 and 2.3. This special case results in a Markovian process, which is not in general true of IVT processes (Barndorff-Nielsen et al., 2014). In fact, the model is similar to the popular Poissonian INAR(1) model, introduced in McKenzie (1985) and Al-Osh and Alzaid (1987). An illustration of the exponential trawl set, A_t, dragged through time, together with a simulation of the resulting Poisson-Exponential IVT trawl process $X_t = L(A_t)$, is seen in Figure 1. The parameters used are $\lambda = 1$ and $\nu = 5$. At each time point t, the value of X_t (bottom plot) is the number of points inside the trawl set A_t (top plot).

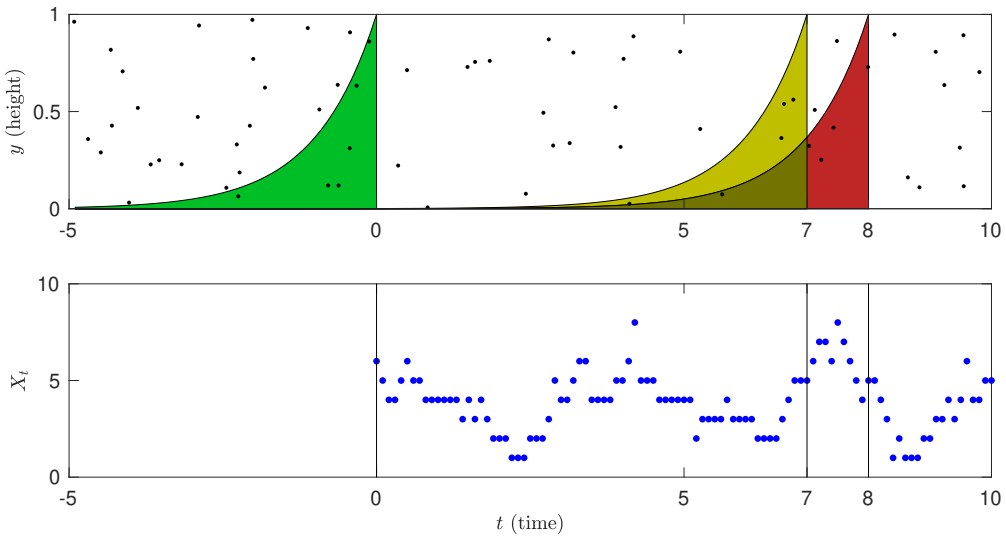

Figure 1: *Top: Simulation of a Poisson Lévy basis on* $\mathbb{R} \times [0,1]$ *(black dots) with an exponential trawl set* A_t *(shaded) superimposed at three periods in time,* $t \in \{0, 7, 8\}$*. Bottom: The associated trawl process* $X_t = L(A_t)$*, given by the number of 'points' inside the trawl set* A_t *at time* t*. The intensity of the Poisson random measure is* $\|\eta\| = \eta(1) = \nu = 5$ *and the parameter controlling the exponential trawl function,* $d(s) = \exp(\lambda s)$*, is* $\lambda = 1$*.*

3 Estimation of integer-valued trawl processes

Barndorff-Nielsen et al. (2014) showed that the parameter vector θ of an IVT process can be consistently estimated using a generalized methods of moments (GMM) procedure. Since we will compare our likelihood-based estimation approach with the method of moments approach, we briefly review this latter approach in Appendix A. In Section 3.1, we develop our likelihood-based approach to estimation of θ.

Both estimation procedures rely on the fact that the IVT process is stationary and mixing. The mixing property of IVT processes obtains from results given in Fuchs and Stelzer (2013), see Barndorff-Nielsen et al. (2014, p. 699). Although mixing in general is sufficient for the consistency of the estimators, the central limit theorem for the likelihood-based estimator (Theorem 3.3 below) relies on the size (or rate) of mixing. We therefore give these details for IVT processes in the following Theorem 3.1. Before we state the result, we recall the definition of α-mixing for a stationary process. Let $\mathcal{F}_{-\infty}^{0} = \sigma(X_t; t \leq 0)$ and, for $m > 0$, $\mathcal{F}_{m}^{\infty} = \sigma(X_t; t \geq m)$, and define the numbers

$$\alpha_m := \sup_{G \in \mathcal{F}_{-\infty}^{0}, H \in \mathcal{F}_{m}^{\infty}} |\mathbb{P}(H \cap G) - \mathbb{P}(H)\mathbb{P}(G)|, \qquad m > 0.$$

The process $X = (X_t)_{t \in \mathbb{R}}$ is α-mixing if $\alpha_m \to 0$ as $m \to \infty$. It is α-mixing of size $-\phi_0$ if $\alpha_m = O(m^{-\phi})$, as $m \to \infty$, for some $\phi > \phi_0$.

Theorem 3.1. *The IVT process X has $\alpha_m = O(\rho(m))$ as $m \to \infty$, where $\rho(m)$ is the autocorrelation function of X.*

Remark 3.1. The autocorrelation functions of the Exponential (Example 2.3) and IG (Example 2.4) IVT models imply that these models are in fact α-mixing with an exponential decay rate. The autocorrelation function of the Gamma (Example 2.5) IVT model implies that it is α-mixing of size $-(H - \epsilon)$ for all $\epsilon > 0$.

3.1 Estimation by composite likelihoods

Due to the non-Markovianity of the IVT process we face computational difficulties when attempting to estimate the model by maximizing the likelihood of the observations: calculation of the full likelihood is overwhelmingly unwieldy in any but the most simple cases (Shephard and Yang, 2016). For this reason, we here propose to use the CL method originally proposed in Lindsay (1988). The main idea behind the CL approach is to specify a quasi-likelihood function which captures the salient features of the data at hand; here this means capturing the features of the Lévy basis, controlling the marginal distribution, and those of the trawl function, controlling the dependence structure. We focus on the so-called pairwise CLs, i.e. we consider the likelihood of the observations of the IVT process using pairs of data points.

3.1.1 Pairwise composite likelihood

Suppose we have $n \in \mathbb{N}$ observations of the IVT process X, $x_1, \dots, x_n$, on an equidistant grid of size $\Delta = T/n$, for some $T > 0$. Define the following likelihood function using pairs of observations k periods apart,

$$CL^{(k)}(\theta; x) := \prod_{i=1}^{n-k} f(x_{i+k}, x_i; \theta), \quad k \geq 1, \tag{3.1}$$

where $f(x_{i+k}, x_i; \theta)$ is the joint probability mass function (PMF) of the observations x_i and x_{i+k}, parametrized by the vector θ. From (3.1), we construct the composite likelihood function

$$\mathcal{L}_{CL}(\theta; x) := \mathcal{L}_{CL}^{(K)}(\theta; x) := \prod_{k=1}^{K} CL^{(k)}(\theta; x) = \prod_{k=1}^{K} \prod_{i=1}^{n-k} f(x_{i+k}, x_i; \theta), \tag{3.2}$$

where $K \in \mathbb{N}$ denotes the number of pairwise likelihoods to include in the calculation of the composite likelihood function. The number K plays the same role as in the case of the GMM estimator (A.1), and it should be chosen such that the salient features of the data are likely to be captured. We give an example of how to select K in the empirical application in Section 6.

Remark 3.2. An alternative composite likelihood function can be constructed by noting that

$$f(x_1, \dots, x_n) = f(x_n | x_{n-1}, \dots, x_1) f(x_{n-1} | x_{n-2}, \dots, x_1) \cdots f(x_2 | x_1) f(x_1).$$

In the Markovian case, this would simplify to

$$f(x_1, \dots, x_n) = f(x_n|x_{n-1})f(x_{n-1}|x_{n-2}) \cdots f(x_2|x_1)f(x_1), \tag{3.3}$$

Equation (3.3) can be viewed as an alternative composite likelihood function. When comparing the performance of the two composite likelihood functions (in (3.2) and (3.3)), we obtained very similar results. Hence we will only present the details for the pairwise likelihood function (3.2) in the following.

Let Θ be a compact parameter space such that the true parameter vector, θ_0, lies in the interior of Θ. The maximum composite likelihood (MCL) estimator of θ is defined as

$$\hat{\theta}^{CL} := \arg\max_{\theta \in \Theta} l_{CL}(\theta; x), \tag{3.4}$$

where $l_{CL}(\theta; x) := \log \mathcal{L}_{CL}(\theta; x)$ is the log composite likelihood function.

To calculate the composite likelihood function $\mathcal{L}_{CL}(\theta; x)$, the individual pairwise PMFs, $f(x_{i+h}, x_i; \theta)$, are required. To derive an expression for these PMFs, note that for two time points $t, u \in \mathbb{R}$, the random variables X_t and X_u can be decomposed as

$$X_t = L(A_t) = L(A_t \backslash A_u) + L(A_t \cap A_u)$$

and

$$X_u = L(A_u) = L(A_u \backslash A_t) + L(A_t \cap A_u),$$

where the random variables $L(A_t \backslash A_u)$, $L(A_u \backslash A_t)$ and $L(A_t \cap A_u)$ are independent. The independence of the three random variables follows by the fact that the Lévy basis L is an independently scattered random measure and that the sets $A_t \backslash A_u$, $A_u \backslash A_t$, and $A_t \cap A_u$ are disjoint. That is, two arbitrary observations of the trawl process, X_t and X_u, can be decomposed into three independent random variables, where one of them, $L(A_t \cap A_u)$, is common between the two observations. Figure 2 illustrates this decomposition, where an exponential trawl (with $\lambda = 1$, see Example 2.3) is plotted at the time points $t = 4$ and $u = 3$ and the sets $A_4 \backslash A_3$, $A_3 \backslash A_4$ and $A_3 \cap A_4$ are indicated. From this discussion, it is immediately clear that the random variables X_t and X_u are independent, conditional on the common part $L(A_t \cap A_u)$.

We can use these insights to derive expressions for the pairwise likelihoods $f(x_{i+k}, x_i; \theta)$ and in Appendix C, we give an infinite sum representation of $f(x_{i+k}, x_i; \theta)$ which is valid in the general integer-valued case. To avoid the practical problems introduced by the infinite sum, Appendix C also presents a simulation-based alternative to estimating $f(x_{i+k}, x_i; \theta)$. Conversely, in the count-valued case, which is our main focus below, the expression for $f(x_{i+k}, x_i; \theta)$ simplifies to a finite sum, as shown in the next proposition. Letting $\mathbb{P}_\theta(B)$ denote the probability of the event B given parameters θ, we have the following.

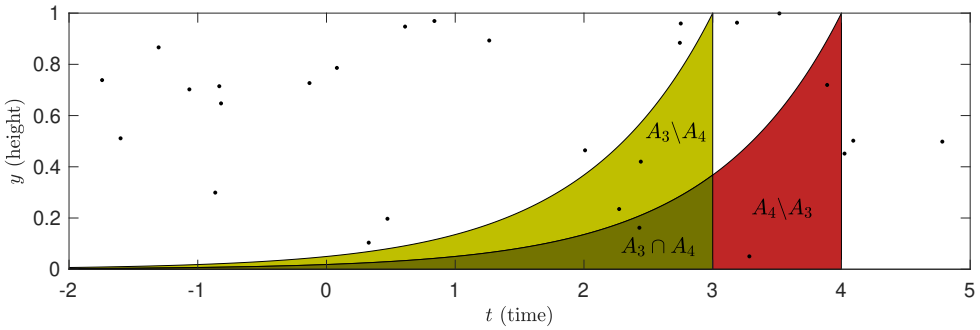

Figure 2: *Decomposition of exponential trawl sets at time $t = 4$, $u = 3$ super-imposed on a Poisson Lévy basis, with $\nu = 5$, on $\mathbb{R} \times [0, 1]$. The exponential trawl parameter is set to $\lambda = 1$.*

Proposition 3.1. *Suppose the Lévy basis L is non-negative, i.e. $\eta(y) = 0$ for $y < 0$. The joint PMF of two observations x_{i+k} and x_i is*

$$f(x_{i+k}, x_i; \theta) = \sum_{c=0}^{\min\{x_{i+k}, x_i\}} \mathbb{P}_\theta \left(L(A_{(i+k)\Delta} \setminus A_{i\Delta}) = x_{i+k} - c\right) \mathbb{P}_\theta \left(L(A_{i\Delta} \setminus A_{(i+k)\Delta}) = x_i - c\right) \cdot \mathbb{P}_\theta \left(L(A_{(i+k)\Delta} \cap A_{i\Delta}) = c\right). \tag{3.5}$$

The probabilities $\mathbb{P}_\theta(\cdot)$ in (3.5) can be expressed as a function of the parameters of the Lévy seed and the trawl function. Indeed, for a Borel set $B \in \mathcal{B}([0, 1] \times \mathbb{R})$ we have

$$\mathbb{P}_\theta(L(B) = x) = \mathbb{P}_\theta(L'_{Leb(B)} = x),$$

where L'_t is a Lévy process with $L'_1 \sim L'$, and L' being the Lévy seed associated to X, see Remark 2.1. Similarly, the Lebesgue measures of the sets $A_{(i+k)\Delta} \setminus A_{i\Delta}$, $A_{i\Delta} \setminus A_{(i+k)\Delta}$, and $A_{(i+k)\Delta} \cap A_{i\Delta}$ can be expressed via the parameters in the trawl function d using Equations (2.7)–(2.9). Indeed,

$$Leb(A_{(i+k)\Delta} \cap A_{i\Delta}) = \int_{-\infty}^{-k\Delta} d(s)ds,$$

and

$$Leb(A_{(i+k)\Delta} \setminus A_{i\Delta}) = Leb(A_{i\Delta} \setminus A_{(i+k)\Delta}) = Leb(A) - Leb(A_{(i+k)\Delta} \cap A_{i\Delta}) = \int_{-k\Delta}^{0} d(s)ds.$$

Plugging these into (3.5) we obtain the pairwise likelihoods, $f(x_{i+k}, x_i; \theta)$, and thus the CL function, $\mathcal{L}_{CL}(\theta; x)$, as a function of θ. The following example illustrates this in the case of the Poisson-Exponential IVT process. Many additional examples are collected in the Supplementary Material.

Example 3.1 (Poisson-Exponential IVT process)**.** Let $L' \sim \text{Poisson}(\nu)$ and $d(s) = \exp(\lambda s)$, $s \leq 0$, for some $\nu, \lambda > 0$. Since $L' \sim \text{Poisson}(\nu)$ we have $L(B) \sim \text{Poisson}(Leb(B)\nu)$ for Borel sets B and hence

$$\mathbb{P}_\theta \left(L(B) = x\right) = (\nu Leb(B))^x e^{-\nu Leb(B)}/x!, \quad x \geq 0.$$

Further, it is not difficult to show that

$$Leb(A_{(i+k)\Delta} \cap A_{i\Delta}) = \lambda^{-1} e^{-\lambda k \Delta} \quad \text{and} \quad Leb(A_{(i+k)\Delta} \setminus A_{i\Delta}) = \lambda^{-1}(1 - e^{-\lambda k \Delta}).$$

Using this, the probabilities in (3.5) can be expressed as a function of ν and λ and hence the maximization (3.4) can be carried out using standard numerical methods.

3.1.2 Asymptotic theory

Because we are only considering dependencies across pairs of observations and not their dependence with the remaining observations, the pairwise composite likelihood function (3.2) can be viewed as a misspecified likelihood. Nonetheless, since the individual PMFs $f(x_{i+k}, x_i; \theta)$ in (3.2) are proper bivariate PMFs, the *composite score function* $\partial l_{CL}(\theta; x)/\partial\theta$ provides unbiased estimating equations and, under certain regularity assumptions, the usual asymptotic results will apply (Cox and Reid, 2004). However, as pointed out in Varin et al. (2011), formally proving the results in the time series case requires a more rigorous treatment. The following three theorems provide the details on the asymptotic theory in the setup of this paper. Detailed mathematical proofs are available in Section 2 of the Supplementary Material Bennedsen et al. (2021). First, we have a Law of Large Numbers.

Theorem 3.2. *Fix $K \in \mathbb{N}$ and let $\hat{\theta}^{CL}$ be given by (3.4) where Θ is a compact space and let the true parameter vector satisfy $\theta_0 \in int(\Theta)$. Then*

$$\hat{\theta}_{CL} \xrightarrow{\mathbb{P}} \theta_0, \quad n \to \infty.$$

The mixing property of IVT processes, presented in Theorem 3.1, implies that the following Central Limit Theorem also holds.

Theorem 3.3. *Let the conditions from Theorem 3.2 hold and assume that the autocorrelation function of the IVT process satisfies $\sum_{n=1}^{\infty} \rho(n) < \infty$. Then,*

$$\sqrt{n}(\hat{\theta}^{CL} - \theta_0) \xrightarrow{d} N\left(0, G(\theta_0)^{-1}\right), \quad n \to \infty,$$

where $G(\theta_0)$ is the Godambe information matrix *(Godambe, 1960) matrix with inverse*

$$G(\theta_0)^{-1} = H(\theta_0)^{-1} V(\theta_0) H(\theta_0)^{-1},$$

where

$$H(\theta_0) = -\sum_{k=1}^{K} \mathbb{E}\left[\frac{\partial^2}{\partial\theta'\partial\theta} \log f(X_{k\Delta}, X_0; \theta)|_{\theta=\theta_0}\right],$$

and

$$V(\theta_0) = \sum_{k=1}^{K} Var\left(\frac{\partial}{\partial\theta} \log f(X_{k\Delta}, X_0; \theta)|_{\theta=\theta_0}\right)$$

$$+2\sum_{k=1}^{K}\sum_{k'=1}^{K}\sum_{i=1}^{\infty} Cov\left(\frac{\partial}{\partial\theta}\log f(X_{k\Delta},X_0;\theta)|_{\theta=\theta_0},\frac{\partial}{\partial\theta'}\log f(X_{(i+k')\Delta},X_{i\Delta};\theta)|_{\theta=\theta_0}\right).$$

Further, the infinite sum in the expression for $V(\theta_0)$ converges.

Theorem 3.3 implies that feasible inference can be conducted using an estimate of the inverse of the Godambe information matrix

$$\hat{G}(\hat{\theta}^{CL})^{-1}=\hat{H}(\hat{\theta}^{CL})^{-1}\hat{V}(\hat{\theta}^{CL})\hat{H}(\hat{\theta}^{CL})^{-1},$$

where $\hat{\theta}^{CL}$ is the maximum composite likelihood estimate from (3.4). Note that while the straightforward estimator $\hat{H}(\hat{\theta}^{CL})=-n^{-1}\frac{\partial}{\partial\theta\partial\theta'}l_{CL}(\hat{\theta}^{CL};x)$ is consistent for $H(\theta)$ due to the stationarity and ergodicity of the IVT process, $\hat{V}(\hat{\theta}^{CL})$ is more difficult to obtain, since the obvious candidate $n^{-1}\frac{\partial}{\partial\theta}l_{CL}(\theta;x)\frac{\partial}{\partial\theta}l_{CL}(\theta;x)'$ vanishes at $\theta=\hat{\theta}^{CL}$, a fact also remarked in Varin and Vidoni (2005). While it is possible to estimate $V(\theta_0)$ using a Newey-West-type kernel estimator (Newey and West, 1987), we obtained more precise results using a simulation-based approach to estimating $V(\theta_0)$. The details of both approaches are provided in Appendix B.

The proof of Theorem 3.3 relies on the IVT process being mixing of size $-\phi_0$ for some $\phi_0>1$. As shown in Theorem 3.1, this excludes IVT processes with very strong memory, e.g. those with autocorrelation function adhering to $\rho(h)=O(h^{-H})$ for $H\in(0,1]$. As mentioned in Remark 3.1, this is for instance the case for the Gamma trawl function (Example 2.5) with $H\in(0,1]$. For these processes, the convergence rate of the MCL estimator $\hat{\theta}^{CL}$ is slower than $\sqrt{n}$, as the following result shows.

Theorem 3.4. *Let the conditions from Theorem 3.2 hold and assume that the autocorrelation function of the IVT process satisfies $\rho(h)=L_\infty(h)h^{-H}$ for some $H\in(0,1]$, where L_∞ is a function which is slowly varying at infinity, i.e. for all $a>0$ it holds that $\lim_{x\to\infty}\frac{L_\infty(ax)}{L_\infty(x)}=1$. Then,*

(i) For all $\epsilon>0$,

$$n^{H/2-\epsilon}(\hat{\theta}^{CL}-\theta_0)\xrightarrow{\mathbb{P}}0,\quad n\to\infty.$$

(ii) Let $J=\dim(\theta_0)$ be the dimension of θ_0 and denote by $\hat{\theta}_i^{CL}$ and $\theta_{0,i}$ the i'th component of the vectors $\hat{\theta}^{CL}$ and θ_0, respectively. Then, for $i=1,2,\ldots,J$,

$$Var\left(n^{H/2+\epsilon}(\hat{\theta}_i^{CL}-\theta_{0,i})\right)\to\infty,\quad n\to\infty.$$

Remark 3.3. Although the CLT in Theorem 3.3 excludes long memory processes, Theorem 3.4 implies that a CLT in the long memory case would have a convergence rate of $n^{H/2}$ for some $H\in(0,1]$. Deriving such a CLT is beyond the scope of the present paper, but we note that related asymptotic results have recently been obtained in Pakkanen et al. (2021), where a non-Gaussian

limit is found. In that paper, the authors derive the asymptotic theory for partial sums of general trawl processes and their Theorem 4 covers the case of integer-valued trawl processes. Based on these findings, we conjecture (and simulation results, not reported here, seem to confirm) that the composite likelihood estimator $\hat{\theta}^{CL}$ has a similar non-Gaussian limit in the long memory case. Investigations of the precise details are left for future work.

3.1.3 Information criteria for model selection

Takeuchi's Information Criterion (Takeuchi, 1976) is an information criterion, which can be used for model selection in the case of misspecified likelihoods. Varin and Vidoni (2005) adapted the ideas of Takuchi to the composite likelihood framework and provided arguments for using the composite likelihood information criterion (CLAIC)

$$CLAIC = l_{LC}(\hat{\theta}^{CL}; x) + \operatorname{tr}\left\{\hat{V}(\hat{\theta}^{CL})\hat{H}(\hat{\theta}^{CL})^{-1}\right\} \tag{3.6}$$

as a basis for model selection, where $\operatorname{tr}\{M\}$ is the trace of the matrix M. Specifically, Varin and Vidoni (2005) suggest picking the model that maximizes $CLAIC$.

To gain some intuition for the CLAIC, let q be the dimension of the parameter vector θ, and note that if instead of the composite likelihood function we had used the correct full likelihood function, the information equality would imply that $VH^{-1} = -I_q$ with I_q being the $q \times q$ identity matrix. Consequently, in this case,

$$CLAIC = l_{CL}(\hat{\theta}^{CL}; x) + \operatorname{tr}\left\{\hat{V}(\hat{\theta}^{CL})\hat{H}(\hat{\theta}^{CL})^{-1}\right\} \approx l_{CL}(\hat{\theta}^{CL}; x) - q,$$

which is the well-known Akaike Information Criterion (AIC, Akaike, 1974). This also explains the notation "CLAIC" for this criterion. Analogous to the usual Bayesian/Schwarz Information Criterion (BIC, Schwarz, 1978), we also suggest the alternative composite likelihood information criterion (Gao and Song, 2010)

$$CLBIC = l_{CL}(\hat{\theta}^{CL}; x) + \frac{\log(n)}{2}\operatorname{tr}\left\{\hat{V}(\hat{\theta}^{CL})\hat{H}(\hat{\theta}^{CL})^{-1}\right\}, \tag{3.7}$$

where n is the number of observations of the data series x. Note that the various models we consider are generally non-nested, whereas most research on model selection using the composite likelihood approach has considered nested model (Ng and Joe, 2014). An analysis of the properties of $CLAIC$ and $CLBIC$ in the non-nested case in the spirit of, e.g., Vuong (1989) would be very valuable but is beyond the scope of the present article.

4 Forecasting integer-valued trawl processes

Let $\mathcal{F}_t = \sigma((X_s)_{s \leq t})$ be the sigma-algebra generated by the history of the IVT process X up until time t and let $h > 0$ be a forecast horizon. The goal of this section is to use the information

available at time t to forecast the future value X_{t+h}. The optimal forecast, in a mean-squared error sense, is the conditional expectation of the future value, i.e. $\tilde{X}_{t+h|t} = \mathbb{E}[X_{t+h}|\mathcal{F}_t]$. However, it is clear that $\tilde{X}_{t+h|t}$ is not *data coherent* in the sense that it is, in general, not integer-valued (Freeland and McCabe, 2004).

Consider instead a probabilistic forecasting approach, where the interest is in the distribution of $X_{t+h}|\mathcal{F}_t$. The advantages of this method are that it can easily generate data coherent point forecasts, e.g. using the median or mode of the distribution, and it provides more information (in the form of the entire predictive distribution) than a simple point forecast. However, since the IVT process X is in general non-Markovian, the distribution of $X_{t+h}|\mathcal{F}_t$ is highly intractable. This problem is similar to the one encountered when considering the likelihood of observations of X, cf. Section 3.1. For this reason, we propose to approximate the distribution of $X_{t+h}|\mathcal{F}_t$ by $X_{t+h}|X_t$, i.e. instead of conditioning on the full information set, we only condition on the most recent observation. Thus, our proposed solution to the forecasting problem is akin to the proposed solution to the problem of the intractability of the full likelihood. That is, instead of considering the full distribution of $X_{t+h}|\mathcal{F}_t$, we use the conditional "pairwise" distribution implied by $X_{t+h}|X_t$.

To fix ideas, let $t \in \mathbb{R}$ and $h > 0$, and consider the random variables $X_t = L(A_t) = L(A_t \cap A_{t+h}) + L(A_t \setminus A_{t+h})$ and $X_{t+h} = L(A_{t+h}) = L(A_t \cap A_{t+h}) + L(A_{t+h} \setminus A_t)$. The goal is to find the conditional distribution of X_{t+h} given X_t. Using similar reasoning as that above, see e.g. Figure 2, it is evident that $L(A_t \cap A_{t+h})$ and $L(A_{t+h} \setminus A_t)$ are independent random variables. Further, since $L(A_{t+h} \setminus A_t)$ is independent of X_t with known distribution, we only need to determine the distribution of $L(A_t \cap A_{t+h})$ given X_t. The following lemma characterises the conditional distribution of $L(A_t \cap A_{t+h})$.

Lemma 4.1. *Let $x \in \mathbb{N} \cup \{0\}$ and $l \in \{0, 1, \dots, x\}$, then*

$$P(L(A_t \cap A_{t+h}) = l | X_t = x) = \frac{\mathbb{P}(L(A_t \setminus A_{t+h}) = x - l)\mathbb{P}(L(A_t \cap A_{t+h}) = l)}{\mathbb{P}(X_t = x)}.$$

The following examples give the details for our two main specifications for the marginal distribution of X_t, the Poisson distribution (Example 2.1) and the Negative Binomial distribution (Example 2.2).

Example 4.1. In the case when $L' \sim \text{Poi}(\nu)$, we get the Binomial distribution:

$$L(A_t \cap A_{t+h})|X_t \sim \text{Bin}\left(X_t, \frac{Leb(A_0 \cap A_h)}{Leb(A_0)}\right).$$

Example 4.2. In the case when $L' \sim \text{NB}(m, p)$, we get the following distribution:

$$\begin{aligned}&\mathbb{P}(L(A_t \cap A_{t+h}) = l | X_t = x) \\ &= \binom{x}{l} \frac{\Gamma(Leb(A_0 \setminus A_h)m + x - l)}{\Gamma(Leb(A_0 \setminus A_h)m)} \frac{\Gamma(Leb(A_0 \cap A_h)m + l)}{\Gamma(Leb(A_0 \cap A_h)m)} \frac{\Gamma(Leb(A_0)m)}{\Gamma(Leb(A_0)m + x)}, \quad x \geq l \geq 0,\end{aligned}$$

where $\binom{x}{l} = \frac{x!}{l!(x-l)!}$ is the binomial coefficient.

Using Lemma 4.1, we can derive the distribution of $X_{t+h}|X_t$, which can be used for probabilistic forecasting. The details for non-negative valued Lévy bases are given in the following proposition.

Proposition 4.1. *Suppose the Lévy basis L is non-negative, i.e. $\eta(y) = 0$ for $y < 0$. Now*

$$\mathbb{P}(X_{t+h} = x_{t+h}|X_t = x_t) = \sum_{c=0}^{\min(x_t, x_{t+h})} \mathbb{P}(L(A_{t+h} \setminus A_t) = x_{t+h} - c)\mathbb{P}(L(A_t \cap A_{t+h}) = c|X_t = x_t).$$

The following corollaries give the specific details for our two main specifications for the marginal distribution of X_t, studied in Examples 4.1 and 4.2 above.

Corollary 4.1. *If $L' \sim \text{Poi}(\nu)$, then*

$$\begin{aligned}
&\mathbb{P}(X_{t+h} = x_{t+h}|X_t = x_t) \\
&= \sum_{c=0}^{\min(x_t, x_{t+h})} \frac{(\nu Leb(A_h \setminus A_0))^{x_{t+h}-c}}{(x_{t+h}-c)!} e^{-\nu Leb(A_h \setminus A_0)} \binom{x_t}{c} \left(\frac{Leb(A_h \cap A_0)}{Leb(A_0)}\right)^c \left(1 - \frac{Leb(A_h \cap A_0)}{Leb(A_0)}\right)^{x_t - c}.
\end{aligned}$$

Corollary 4.2. *If $L' \sim \text{NB}(m, p)$, then*

$$\begin{aligned}
&\mathbb{P}(X_{t+h} = x_{t+h}|X_t = x_t) \\
&= \sum_{c=0}^{\min(x_t, x_{t+h})} (1-p)^{Leb(A_h \setminus A_0)m} p^{x_{t+h}-c} \binom{x_t}{c} \frac{1}{(x_{t+h}-c)!} \\
&\cdot \frac{\Gamma(Leb(A_h \setminus A_0)m + x_{t+h} - c)}{\Gamma(Leb(A_h \setminus A_0)m)} \frac{\Gamma(Leb(A_h \setminus A_0)m + x_t - c)}{\Gamma(Leb(A_h \setminus A_0)m)} \frac{\Gamma(Leb(A_h \cap A_0)m + c)}{\Gamma(Leb(A_h \cap A_0)m)} \frac{\Gamma(Leb(A_0)m)}{\Gamma(Leb(A_0)m + x_t)}.
\end{aligned}$$

If the parameters of an IVT process X with Poisson or Negative Binomial marginal distribution are known, we can use Corollary 4.1 or 4.2, and the calculations for the Lebesgue measures of the trawl sets given in Equations (2.7)–(2.9), for computing the predictive PMFs and thus for forecasting. When the true parameter values are unknown, they can be estimated using the MCL estimator suggested above, and plugged into the formulas to arrive at estimates of the predictive PMFs.

5 Monte Carlo simulation experiments

In a simulation study, we examine the finite sample properties of the composite likelihood-based estimation procedure of Section 3.1; the details are given in Section 5.1 below. We also conducted a simulation study to assess the finite sample properties of the model selection procedure introduced in Section 3.1.3; the results and details of this study are reported in the Supplementary Material.

The IVT framework is very flexible and there are many possible choices of data generating processes (DGPs) to use in the simulation studies. Here, we will consider the six combinations of the two marginal distributions, given in Examples 2.1 and 2.2, and three correlations structures, given in Examples 2.3, 2.4, and 2.5. In other words, we consider the Poisson-Exponential

Table 1: *Parameter values used in simulation studies*

DGP	ν	m	p	λ	δ	γ	H	α
P-Exp	17.50			1.80				
P-IG	17.50				1.80	0.80		
P-Gamma	17.50						1.70	0.80
NB-Exp		7.50	0.70	1.80				
NB-IG		7.50	0.70		1.80	0.80		
NB-Gamma		7.50	0.70				1.70	0.80

Parameter values for the six different DGPs used in the simulation studies of Section 5. See Examples 2.1, 2.2, 2.3, 2.4, and 2.5 for details. The value $\nu = mp/(1-p)$ with $m = 7.5$ and $p = 0.70$ is chosen such that the first moment of the Poisson and Negative Binomial Lévy bases are matched. Marginal distributions and autocorrelation functions implied by these parameter values are shown in Figure 3.

(P-Exp), the Poisson-Inverse Gaussian (P-IG), the Poisson-Gamma (P-Gamma), the Negative Binomial-Exponential (NB-Exp), the Negative Binomial-Inverse Gaussian (NB-IG), and the Negative Binomial-Gamma (NB-Gamma) IVT models. Note that the first model contains two free parameters, the second, third, and fourth models three free parameters, and the fifth and sixth models four free parameters. Since the Lévy bases considered here are non-negative-valued, we will use Proposition 3.1 for calculation of the pairwise likelihoods.

The parameter values used in the simulation studies are given in Table 1 and the implied marginal distributions and autocorrelation structures are shown in Figure 3. The figure illustrates the difference between the six DGPs: those based on the Poisson Lévy basis have a more concentrated marginal distribution compared to those based on the Negative Binomial Lévy basis; those based on the exponential trawl function have smaller degrees of autocorrelation (memory) than those based on the Inverse Gaussian trawl function, and the Gamma trawl function can exhibit still greater autocorrelation.

The choice of parameter values used in the simulation studies below and given in Table 1 are based on the estimates obtained in the empirical study in Section 6. We have found the finite sample properties of the methods proposed in this paper to be relatively robust to the exact choice of parameter values; in the Supplementary Material we illustrate this, by performing simulation studies similar to the ones presented here, but with different settings for the true parameter values.

5.1 Finite sample properties of the MCL estimator

Consider n equidistant observations of an IVT process on an equidistant grid of size 0.10, i.e. $X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$ with $\Delta = 0.10$. We simulate 500 Monte Carlo replications of such time series and in each iteration estimate the parameters of the model using the MCL approach of Equation (3.4). For the IVT models based on the exponential trawl function (Example 2.3), we set $K = 1$, while we set $K = 10$ for the remaining IVT models.[3] In extensive simulation experiments (not

[3]Our analyses have shown that $K = 1$ will deliver good estimation results for the IVT models with exponential trawl functions, but poor estimation results for the models with more other trawl functions. This is not surprising

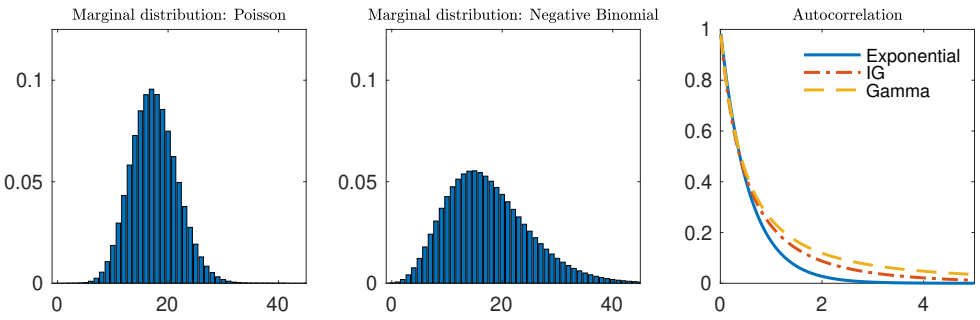

Figure 3: *Marginal distributions of the Lévy bases and autocorrelations of the DGPs used in the simulation studies of Section 5. The marginal distribution and autocorrelation structure of IVT processes can be specified independently, resulting in six different DGPs in this setup (P-Exp, P-IG, P-Gamma, NB-Exp, NB-IG, NB-Gamma). See Examples 2.1, 2.2, 2.3, 2.4, and 2.5 for details. The parameter values used to produce the plots are given in Table 1.*

reported here), we verified that the results are robust to the choice of K. In the Supplementary Material (Tables S2–S7), we report the median, the median bias, and the root median squared error (RMSE) of the estimator, calculated over the 500 Monte Carlo replications for various values of n. The reason for reporting the median, instead of the mean, is that we found that when the number of observations, n, is small, the estimation approach will occasionally result in large outliers in few of the Monte Carlo runs, thus skewing the results (this was the case for both the MCL and GMM estimators).

As mentioned, previous applied work using IVTs have mainly relied on the moment-based estimator. We therefore compare the GMM estimation procedure, laid forth in Section A, with the MCL estimator suggested in this paper. Figure 4 plots the RMSE of the MCL estimator of a given parameter divided by the RMSE of the GMM estimator of the same parameter for the six DGPs of Table 1. Thus, numbers smaller than one indicate that the MCL estimator has lower RMSE than the GMM estimator and vice versa for numbers larger than one. We see that for most parameters in most of the DGPs, the MCL estimator outperforms the GMM estimator substantially; indeed, in many cases, the RMSE of the MCL estimator is around 50% that of the GMM estimator. The exception seems to be the trawl parameters, i.e. the parameters controlling the autocorrelation structure, in the case of the Gamma and IG trawls, where the GMM estimator occasionally performs on par with the MCL estimator. However, in most cases, it appears that the MCL estimator is able to provide large improvements over the GMM estimator.

since the correlation structure for an IVT model with an exponential trawl is very simple, while it is more complicated for other IVT processes. The upshot is that choosing $K = 1$ is sufficient for the simple exponential trawl-based IVT models, while it is necessary to choose $K > 1$ to obtain good results for IVT models constructed using other trawl functions. This is analogous to the situation for the GMM estimator, where the estimator of the λ parameter in the exponential trawl function has a closed-form solution using only the autocorrelation function calculated at the first lag, cf. Appendix A.

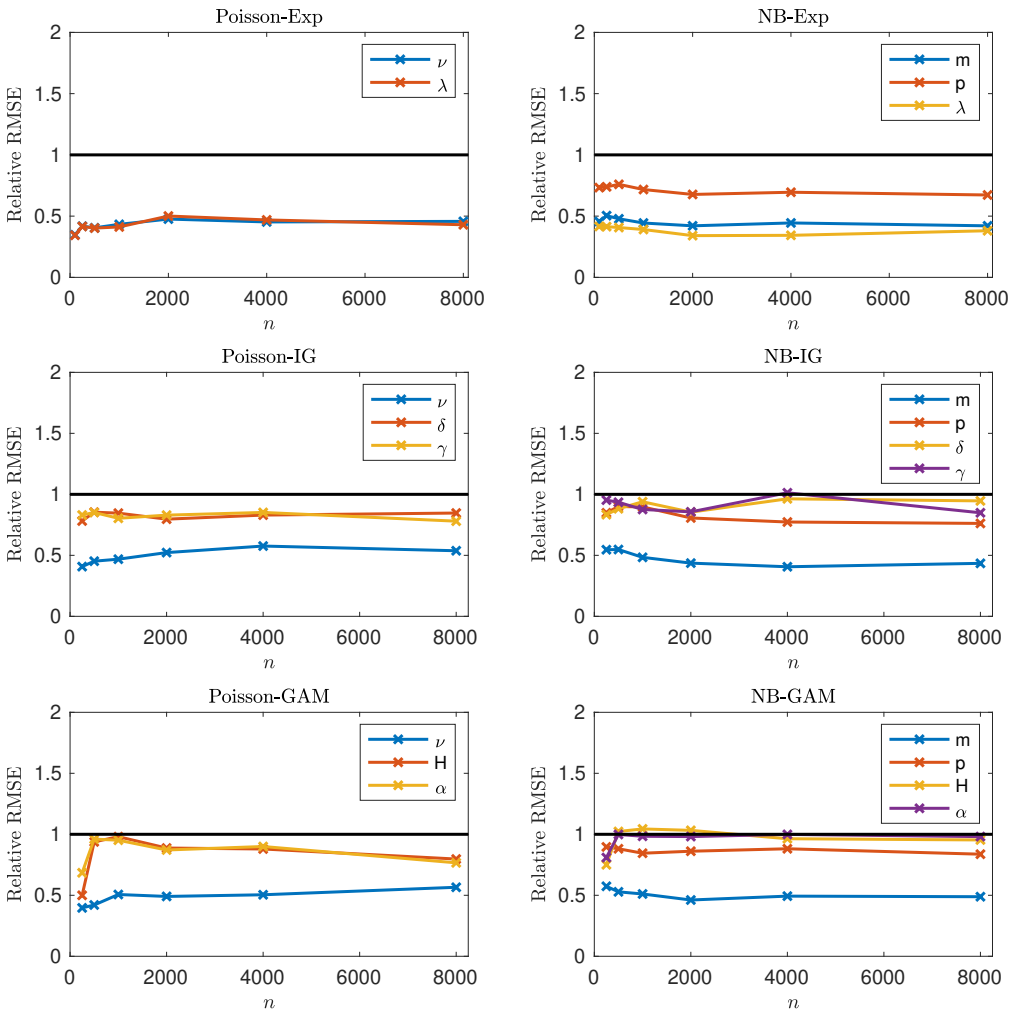

Figure 4: *Root median square error (RMSE) of the MCL estimator* (3.4) *divided by the RMSE of the GMM estimator. The underlying IVT process* X_t *is simulated on the grid* $t = \Delta, 2\Delta, \ldots, n\Delta$*, with* $\Delta = 0.10$*, see Table 1 for the values of the parameters used in the simulations. For the Poisson-Exp and NB-Exp models we set* $K = 1$ *in* (A.1) *and* (3.4)*; for the other models we set* $K = 10$*.*

6 Application to financial bid-ask spread data

In this section, we apply the IVT modelling framework to the bid-ask spread of equity prices. The bid-ask spread has been extensively studied in the market microstructure literature, see, e.g., Huang and Stoll (1997) and Bollen et al. (2004). An application similar to the one studied in this section was considered in Barndorff-Nielsen et al. (2014).

To illustrate the use of methods proposed in this paper, we study the time series of the bid-ask spread, measured in U.S. dollar cents, of the Agilent Technologies Inc. stock (ticker: A) on a single day, May 4, 2020. The A stock is traded on the New York Stock Exchange, which is open from 9.30AM to 4PM. To avoid opening effects, we consider the data from 10:30AM to 4PM,

i.e. we discard the first 60 minutes of the day. Our data is gathered from the Trade and Quote database and cleaned using the approach proposed in Barndorff-Nielsen et al. (2009). The data is available at a very high frequency but to obtain equidistant data, we sample the observations with $\Delta = \frac{1}{12}$ minutes (i.e. 5 seconds) time steps, using the previous tick approach, starting at 10:30AM, resulting in $n = 3961$ observations.

Let s_t be the bid-ask spread level at time t, the time series of which is displayed in the top panel of Figure 5. Since the minimum spread level in the data is one tick (one dollar cent), we work on this time series minus one, i.e on $x_t = s_t - 1$. Based on visual inspection of the empirical autocorrelation (for choosing the trawl function, i.e. for specifying the autocorrelation structure of the model) and the histogram of the spread level (for choosing the Lévy basis, i.e. for specifying the marginal distribution), Barndorff-Nielsen et al. (2014) chose the NB-IG IVT model for data similar to those studied here. Using the methods proposed in this paper, we can now approach the model selection with more rigour. We first inspect the empirical autocorrelation of the data (shown in the right panels of Figure 5) which shows evidence of a very persistent process; we therefore set K to a moderately large value to accurately capture the dependence structure of the data. Here, we choose $K = 10$ but the results are robust to other choices. For instance, we experimented with $K = 20$ with comparable results.

Using this setting, we calculated the maximized composite likelihood value, CL, and the two information criteria, $CLAIC$ and $CLBIC$, obtained for these data using the six models considered in Section 5. The results are shown in Table 2. The table shows that the NB-Gamma model is the preferred model on all three criteria, while the second-best model is the NB-IG model.

To further examine the fit of the various models, the bottom six rows of Figure 5 plots the empirical autocorrelation (left; blue bars) and the empirical marginal distribution of the spread level (right; blue bars). Each respective row also shows the fit of one of the six models considered in Table 2; the parameter estimates corresponding to the models are given in Table 3. The fit of the models shown in the bottom six panels of Figure 5 and the selection criteria of Table 2, indicate that the models based on the Negative Binomial distribution are preferred to the models based on the Poisson distribution. We conclude that, for this data series, the Poisson distribution is unable to accurately describe the marginal distribution of the spread level sampled every 5 seconds. That the Gamma and IG trawl functions are preferred to the Exponential trawl function indicates that the Exponential autocorrelation function is not flexible enough to capture the correlation structure of the data. By both visual inspection of the autocorrelations in Figure 5 and the selection criteria of Table 2, we conclude that the NB-Gamma model is the preferred model for these data. As shown in Table 3, this model has $\hat{H} = 1.70$ (s.e. 0.74), implying that the model possesses a slowly decaying autocorrelation structure, albeit not the long memory property.

Table 2: *Model selection results: bid-ask spread data*

Model:	Poisson-Exponential	Poisson-IG	Poisson-Gamma	NB-Exponential	NB-IG	NB-Gamma
CL	-244125.5	-242885.2	-242835.8	-216363.9	-216318.1	$\mathbf{-216313.5}$
$CLAIC$	-244125.7	-242899.1	-242855.8	-216364.0	-216318.5	$\mathbf{-216313.9}$
$CLBIC$	-244126.4	-242942.8	-242918.8	-216364.1	-216319.7	$\mathbf{-216315.1}$

Composite likelihood and information criteria values for fitting the A bid-ask spread data on May 4, 2020, shown in the top plot of Figure 5, calculated using six different models as given in the top row of the table using $K = 10$. The maximum value for a given criteria (i.e. row-wise) is given in bold. The parameter estimates corresponding to the fits are given in Table 3.

Table 3: *Estimation results: bid-ask spread data*

DGP	ν	m	p	λ	δ	γ	H	α
P-Exp	28.9319			4.0399				
	(0.6644)			(0.0904)				
P-IG	1444.9				0.0015	0.0000		
	(236.1378)				(0.0336)	(0.0013)		
P-Gamma	98.1513						0.5491	0.0400
	(7.5002)						(0.0484)	(0.0059)
NB-Exp		6.4273	0.6665	1.7835				
		(0.9324)	(0.0215)	(0.1349)				
NB-IG		7.7104	0.6675		1.7816	0.8292		
		(1.0111)	(0.0238)		(0.4436)	(0.2094)		
NB-Gamma		7.7336	0.6675				1.7020	0.7897
		(1.1316)	(0.0260)				(0.7365)	(0.3363)

Parameter estimates (standard errors in parentheses) from the six different DGPs when applied to the bid-ask spread data of A on May 4, 2020 using the MCL estimator with $K = 10$. The standard errors have been obtained using the simulation-based approach to estimating the asymptotic covariance matrix of the MCL estimator, see Appendix B. See Figure 5 for the resulting fits of the models to the empirical distribution and autocorrelation.

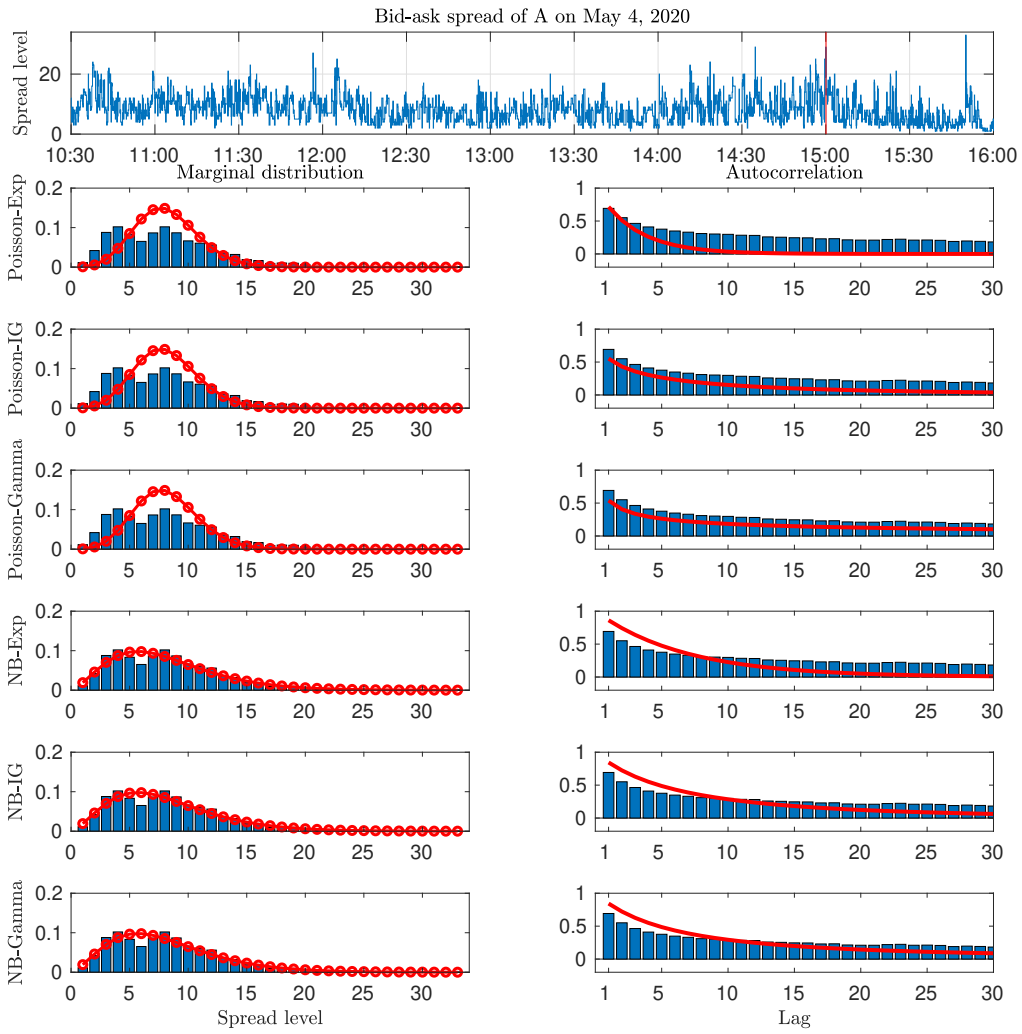

Figure 5: *Analysis of the A spread level on May 4, 2020. Top: The data (spread level in cents) from* 10:30*AM to* 4:00*PM sampled every* 5 *seconds. The vertical red line separates the initial in-sample period (left) from the out-of-sampe period (right), used in the forecasting exercise in Section 6.1. The bottom six rows show the empirical autocorrelation (left; blue bars) and the empirical marginal distribution of the spread level (right; blue bars) together with the fits from the six IVT models (red lines). The parameters of the models have been estimated using the MCL estimator* (3.4) *with* $K = 10$*, see Table 3.*

6.1 Forecasting the spread level

This section illustrates the use of IVT models for forecasting, as outlined in Section 4. The aim is to forecast the future spread level of the A stock on May 4, i.e. the data studied above and plotted in the top panel of Figure 5. We set aside the first $n_1 = 3221$ observations as an "in-sample period" for initial estimation of the parameters of the models, see the vertical red line in the top panel of Figure 5 for the placement of this split. We then forecast the spread level from 5 seconds until 100 seconds into the future, using the approach presented in Section 4. That is, we forecast

$x_{n_1+1}, x_{n_1+2}, \ldots, x_{n_1+20}$ given the current value x_{n_1}. After this, we update the in-sample data set with one additional observation so that this sample now contains $n_2 = n_1 + 1 = 3222$ observations. Then we again forecast the next 20 observations, $x_{n_2+1}, x_{n_2+2}, \ldots, x_{n_2+20}$, given x_{n_2}. We repeat this procedure until the end of the sample, which yields $n_{oos} = n - n_1 - 20 = 720$ out-of-sample forecasts for each forecast horizon. To ease the computational burden, we only re-estimate the model every 24 periods (i.e. every 2 minutes).

To evaluate the forecasts, we consider four different loss metrics. The first two, the mean absolute error (MAE) and the mean squared error (MSE), are often used in econometric forecasting studies of real-valued data (e.g. Elliott and Timmermann, 2016). For a forecast horizon $h = 1, 2, \ldots, 20$, define the mean absolute forecast error,

$$MAE(h) = \frac{1}{n_{oos}} \sum_{i=n_1+h}^{n-(20-h)} |x_i - \hat{x}_{i|i-h}|,$$

where $\hat{x}_{i|i-h}$ is the h-step ahead forecast of x_i, constructed using the information available up to observation $i - h$. That is, $\hat{x}_{i|i-h}$ is the point forecast of x_i coming from a particular IVT model, such as the conditional mean, median, or mode. In what follows, we set $\hat{x}_{i|i-h}$ equal to the estimated conditional mode, i.e. $\hat{x}_{i|i-h} = \arg\max_k \widehat{\mathbb{P}}(X_{i|i-h} = k)$, where $\widehat{\mathbb{P}}$ is the estimated predictive PMF of the IVT model, see Section 4.[4] Define also the mean squared forecast error

$$MSE(h) = \frac{1}{n_{oos}} \sum_{i=n_1+h}^{n-(20-h)} (x_i - \hat{x}_{i|i-h})^2.$$

We consider two additional loss metrics, designed to directly evaluate the accuracy of the estimated predictive PMF $\widehat{\mathbb{P}}$, which is arguably more relevant to the problem at hand than MAE and MSE. The first is the logarithmic score (Elliott and Timmermann, 2016, p. 30):

$$logS(h) = \frac{1}{n_{oos}} \sum_{i=n_1+h}^{n-(20-h)} -\log \widehat{\mathbb{P}}(X_{i|i-h} = x_i),$$

where x_i is the realized outcome. The second is the ranked probability score (RPS; Epstein, 1969):

$$RPS(h) = \frac{1}{n_{oos}} \sum_{i=n_1+h}^{n-(20-h)} \sum_{k=0}^{M} (\widehat{F}_{i|i-h}(k) - \mathbb{I}_{\{x_i \leq k\}})^2,$$

where $\widehat{F}_{i|i-h}(k) = \sum_{j=0}^{k} \widehat{\mathbb{P}}(X_{i|i-h} = j)$ is the estimated cumulative distribution function of $X_i|x_{i-h}$ coming from a given model and $\mathbb{I}_{\{x_i \leq k\}}$ is the indicator function of the event $\{x_i \leq k\}$.

Figure S8 shows the four different forecast loss metrics for the preferred NB-Gamma IVT model as a ratio of the forecasting loss of a benchmark model in the out-of-sample forecasting exercise described above. The numbers plotted in the figure are $Loss(h)_{NB-Gamma}/Loss(h)_{benchmark}$, where

[4] We also experimented with specifying $\hat{x}_{i|i-h}$ as the estimated conditional mean, and found similar results to those reported below. These results can be found in Figures S6–S8 of the Supplementary Material file.

"*Loss*" denotes one of the four loss metrics given above and $h = 1, 2, \ldots, 20$ denotes the forecasting horizon. Thus, numbers less than one favour the NB-Gamma model compared to the benchmark model and vice versa for numbers greater than one. Initially, we choose the Poisson-Exponential IVT process as benchmark model; as remarked above, this process is identical to the Poissonian INAR(1) model, which is often used for forecasting count-valued data (e.g. Freeland and McCabe, 2004; McCabe and Martin, 2005; Silva et al., 2009). It is evident from Figure S8 that losses from the NB-Gamma model are smaller than those from the benchmark model for almost all forecast horizons and loss metrics, the only exception being the short forecast horizons in case of MSE loss.[5] In the case of the the two most relevant loss functions for evaluating the predictive distribution, the logS and RPS, the reduction in losses are substantial for all forecast horizon, on the order of 20%.

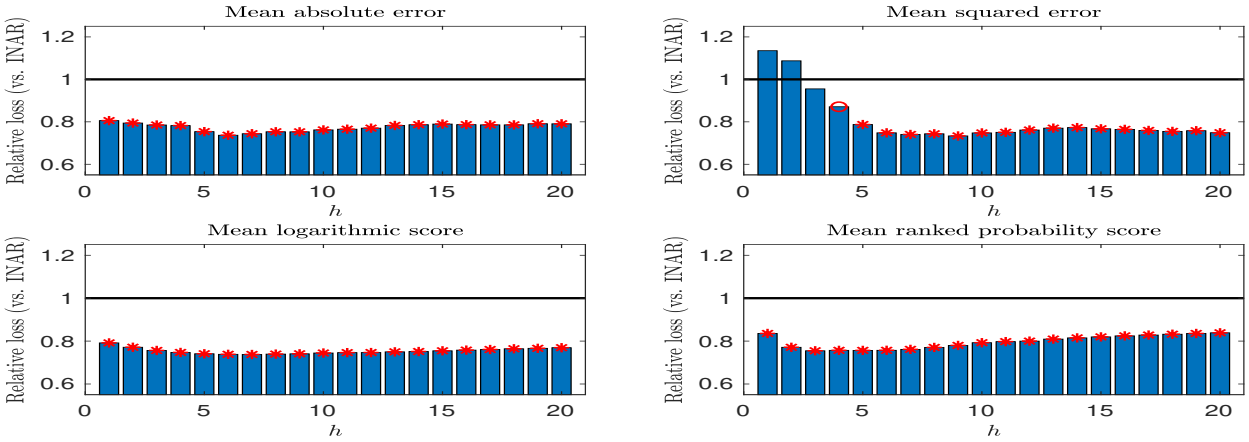

Figure 6: *Forecasting the spread level of the A stock on May 4, 2020. Four different loss metrics and twenty forecast horizons,* $h = 1, 2, \ldots, 20$. *The numbers plotted are relative average losses of the NB-Gamma forecasting model, cf. Section 4, compared with the Poissonian INAR(1) model, over* $n_{oos} = 720$ *out-of-sample forecasts. (The vertical red line in the top panel of Figure 5 denotes the beginning of the out-of-sample period.) A circle above the bars indicates rejection null of equal forecasting performance between the two models, against the alternative that the NB-IG model provides superior forecasts, using the Diebold-Mariano (Diebold and Mariano, 1995) test at a* 5% *level; an asterisk denotes rejection at a* 1% *level.*

To assess whether these loss-differences are also statistically significant, we perform the Diebold-Mariano test of superior predictive ability (Diebold and Mariano, 1995). The null hypothesis of the statistical test is that the two models have equal predictive power, while the alternative hypothesis is that the NB-Gamma model provides superior forecasts compared to the benchmark model. In Figure S8, a circle (asterisk) denotes rejection of the Diebold-Mariano test at a 5% (1%) level.

[5]The poor relative MSE numbers of the NB-Gamma model for the short forecast horizon are caused by our choice of using the conditional mode as a point forecast. Indeed, if we instead use the conditional mean as the point forecast, the NB-Gamma model outperforms the benchmark model in terms of MSE also for the shorter forecast horizons (Supplementary Material, Figure S6).

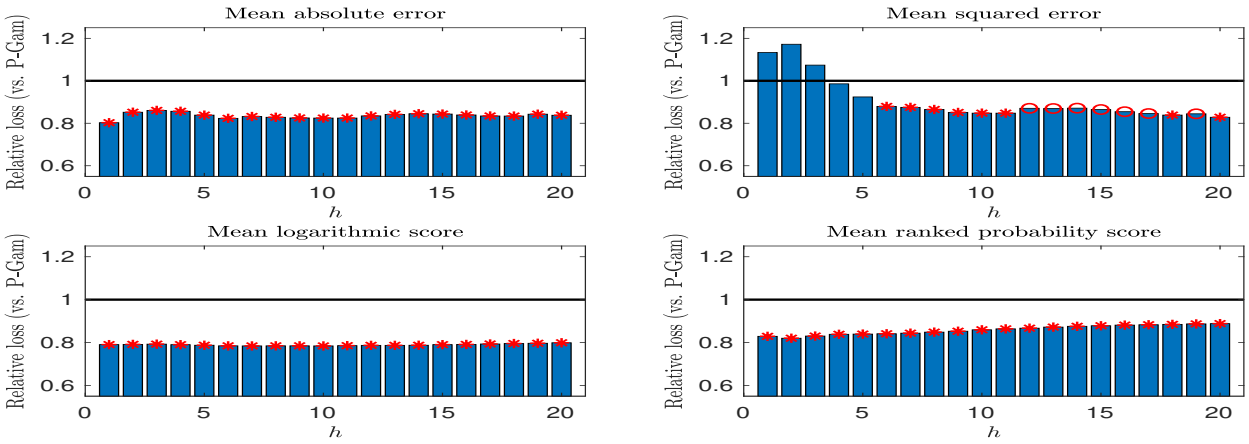

Figure 7: *Forecasting the spread level of the A stock on May 4, 2020. Four different loss metrics and twenty forecast horizons,* $h = 1, 2, \ldots, 20$. *The numbers plotted are relative average losses of the NB-Gamma forecasting model, cf. Section 4, compared with the Poisson-Gamma model, over* $n_{oos} = 720$ *out-of-sample forecasts. (The vertical red line in the top panel of Figure 5 denotes the beginning of the out-of-sample period.) A circle above the bars indicates rejection null of equal forecasting performance between the two models, against the alternative that the NB-IG model provides superior forecasts, using the Diebold-Mariano (Diebold and Mariano, 1995) test at a* 5% *level; an asterisk denotes rejection at a* 1% *level.*

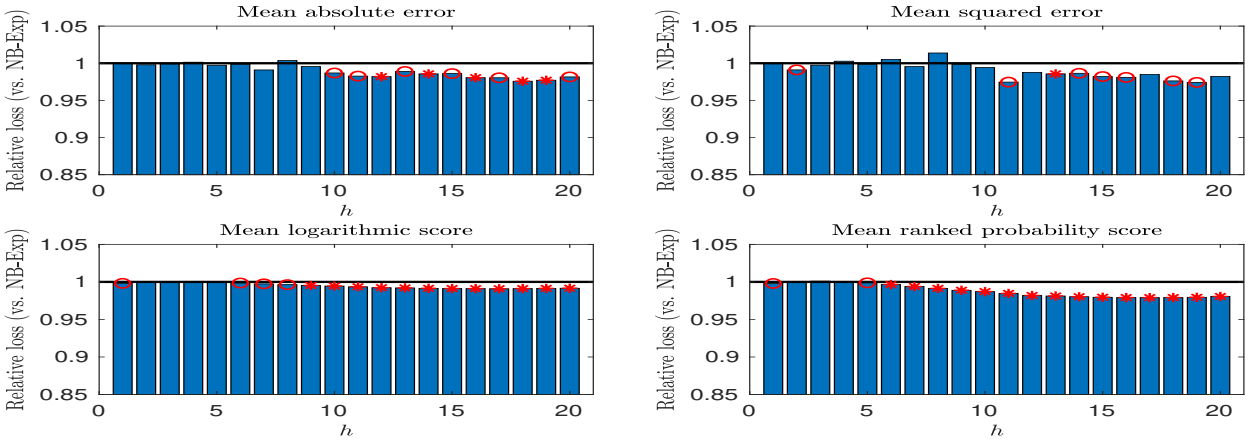

Figure 8: *Forecasting the spread level of the A stock on May 4, 2020. Four different loss metrics and twenty forecast horizons,* $h = 1, 2, \ldots, 20$. *The numbers plotted are relative average losses of the NB-Gamma forecasting model, cf. Section 4, compared with the NB-Exponential model, over* $n_{oos} = 720$ *out-of-sample forecasts. (The vertical red line in the top panel of Figure 5 denotes the beginning of the out-of-sample period.) A circle above the bars indicates rejection null of equal forecasting performance between the two models, against the alternative that the NB-IG model provides superior forecasts, using the Diebold-Mariano (Diebold and Mariano, 1995) test at a* 5% *level; an asterisk denotes rejection at a* 1% *level.*

The test rejects the null hypothesis of equal predictive ability for almost all forecast horizons and loss metrics at a 1% level. These results indicate that there are potentially significant gains to be had from using the IVT framework for forecasting count-valued data such as those studied in this section.

To investigate whether the increased forecast performance of the NB-Gamma model comes from having a more flexible marginal distribution than the Poissonian INAR(1) benchmark model (Negative Binomial vs. Poisson marginals) or from having a more flexible correlation structure (polynomial decay vs. exponential decay) or both, we compare the forecasts from NB-Gamma model to those coming from a P-Gamma model and from a NB-Exp model. These results are given in Figures S6 and S7, respectively. From Figure S6, we see that the NB-Gamma model outperforms the P-Gamma model considerably, indicating that it is important to use a model with Negative Binomial marginals when forecasting these data. Again, the exception is the short forecast horizons in the case of MSE loss. From Figure S7, we see that for the shorter forecast horizons, the NB-Exp model performs on par with the NB-Gamma model, but for the longer forecast horizons, the NB-Gamma model is superior. Hence, when forecasting, it also appears to be important to specify a model with an accurate autocorrelation structure, especially for longer forecasting horizons.

7 Conclusion

This paper has developed likelihood-based methods for estimation, inference, model selection, and forecasting of IVT processes. Since the full likelihood of IVT processes is intractable, we propose to estimate the parameters of the model by maximizing the pairwise likelihood of the data. We proved consistency and asymptotic normality of this MCL estimator and provided the details on how to conduct feasible inference and model selection. We also developed a pairwise approach to approximating the conditional predictive PMF of the IVT process, which can be used for forecasting integer-valued data. All these methods are implemented in a freely available software package written in the MATLAB programming language.

In a simulation exercise, we demonstrated the good properties of the MCL estimator compared to the often-used method of moments-based estimator. Indeed, the reduction in root median squared error of the MCL estimator was in many cases more than 50% compared to the corresponding GMM estimator.

In an empirical application to financial bid-ask spread data, we illustrated the model selection procedure and found that the Negative Binomial-Gamma IVT model provided the best fit to the data. Using the forecast tools developed in the paper, we saw that this model outperformed the often-used Poissonian INAR(1) model considerably, resulting in a reduction in forecast loss on the order of 20% for most forecast horizons. We demonstrated that most of the superior forecasting performance came from accurate modelling of the marginal distribution of the data; however,

we also found that it was beneficial to carefully model the autocorrelation structure, especially for longer forecasting horizons. These findings highlight the strengths of the IVT modelling framework, where the marginal distribution and autocorrelation structure can be modelled independently in a very flexible fashion.

A Method of moments-based estimation of IVT processes

For a parametric IVT model, let the parameter vector θ of the model be given by $\theta = (\theta_d, \theta_L)$, where θ_d contains the parameters governing the trawl function d and θ_L contains the parameters governing the marginal distribution of the IVT process, as specified by the underlying Lévy seed L'. For instance, in the case of the Poisson-Exponential IVT process considered above, cf. Figure 1, we would have $\theta_d = \lambda$ and $\theta_L = \nu$.

Because the correlation structure of an IVT process is decoupled from its marginal distribution, the theoretical autocorrelation function of the process will not depend on θ_L, and we can thus estimate θ_d in a first step, using the empirical autocorrelations of the data. To be precise, let $\rho_{\theta_d}(h)$ be the parametric autocorrelation function as implied by the trawl function d, see Equation (2.10), and let $\hat{\rho}(k)$ be the estimate of the empirical autocorrelation of the data at lag k. The GMM estimator of θ_d is

$$\hat{\theta}_d^{GMM} := \arg\min_{\theta_d \in \Theta_d} \sum_{k=1}^{K} \left(\rho_{\theta_d}(k) - \hat{\rho}(k)\right)^2, \tag{A.1}$$

where $K \geq 1$ denotes the amount of lags to include in the estimation and Θ_d is the parameter space of the trawl parameters in θ_d.[6]

For the estimation of the parameters θ_L governing the marginal distribution of the IVT process, recall that the j'th cumulant κ_j of X_t is given by

$$\kappa_j = Leb(A) \cdot \kappa_j', \quad j = 1, 2, \ldots,$$

where κ_j' is the j'th cumulant of the Lévy seed L'. Using the estimates $\hat{\theta}_d^{GMM}$ from the first step, we can estimate the Lebesgue measure of the trawl set as

$$\widehat{Leb(A)} = \int_0^\infty \hat{d}(-s)ds, \tag{A.2}$$

where $\hat{d}(\cdot)$ denotes the estimate of the trawl function implied by the estimated trawl parameters $\hat{\theta}_d^{GMM}$. The parameters governing the marginal distribution, θ_L, can now be estimated as follows. Let r be the number of elements in θ_L and denote by $\widehat{Leb(A)}$ the estimate of the Lebesgue measure

[6] In the case of the IVT process with an exponential trawl function $d(s) = \exp(\lambda s)$, $s \leq 0$, we use a closed-form estimator of the λ parameter using only the autocorrelation function calculated at the first lag, that is $\hat{\lambda}^{GMM} = -\log\hat{\rho}(1)/\Delta$, where $\Delta > 0$ is the equidistant time between observations.

obtained from (A.2). Estimates of the cumulants, $\hat{\kappa}_j$, can be obtained straightforwardly by calculating the empirical cumulants of the data. Let I_r be a set of r distinct natural numbers (e.g., the numbers from 1 to r). Now

$$\hat{\kappa}_j = \widehat{Leb(A)} \cdot \kappa'_j, \quad j \in I_r,$$

defines r equations in the r unknowns θ_L. GMM estimates of the elements in θ_L, $\hat{\theta}_L^{GMM}$, can be obtained by solving these r equations. Finally, set $\hat{\theta}^{GMM} := (\hat{\theta}_d^{GMM}, \hat{\theta}_L^{GMM})$, which is the method of moments-based estimator of θ.

B Practical details on feasible inference using the MCL estimator

As shown in Theorem 3.3 of Section 3.1.2, the asymptotic variance of the maximum composite likelihood estimator, $\hat{\theta}_{CL}$ is given by the inverse Godambe information matrix,

$$G(\theta_0)^{-1} = H(\theta_0)^{-1} V(\theta_0) H(\theta_0)^{-1}.$$

As mentioned, the matrices $H(\theta_0)$ and $V(\theta_0)$ can be consistently estimated by

$$\begin{aligned}
\hat{H}(\hat{\theta}^{CL}) &= -\frac{1}{n} \frac{\partial}{\partial\theta\partial\theta'} l_{CL}(\hat{\theta}^{CL}; x) \\
&= -\frac{1}{n} \sum_{k=1}^{K} \sum_{i=1}^{n-k} \frac{\partial^2}{\partial\theta\theta'} \log f(x_{(i+k)\Delta}, x_{i\Delta}; \theta)|_{\theta=\hat{\theta}^{CL}}, \\
\hat{V}_{HAC}(\hat{\theta}^{CL}) &= \hat{\Sigma}_0 + \sum_{j=1}^{q} \left(1 - \frac{j}{q+1}\right) (\hat{\Sigma}_j + \hat{\Sigma}'_j),
\end{aligned}$$

where

$$\hat{\Sigma}_j := \frac{1}{n} \sum_{k=1}^{K} \sum_{k'=1}^{K} \sum_{i=1}^{n-j-k'} \frac{\partial}{\partial\theta} \log f(x_{(i+k)\Delta}, x_{i\Delta}; \theta)|_{\theta=\hat{\theta}^{CL}} \frac{\partial}{\partial\theta'} \log f(x_{(i+j+k')\Delta}, x_{(i+j)\Delta}; \theta))|_{\theta=\hat{\theta}^{CL}},$$

and $q \in \mathbb{N}$ is the number of autocorrelation terms to take into account in the HAC estimator. The Hessian, $H(\theta_0)$, is straightforwardly estimated by the above expression. Indeed, a numerical approximation of this matrix is often directly available as output from the software maximizing the composite likelihood function. We have found that while this estimator $\hat{H}(\hat{\theta}^{CL})$ is quite precise, the HAC estimator $\hat{V}_{HAC}(\hat{\theta}^{CL})$ can be rather imprecise. In practice, we therefore recommend estimating $V(\theta_0)$ using simulation-based approach; the details are given in the following Appendix B.1.

B.1 Simulation-based approach to estimating the asymptotic covariance matrix

To obtain a simulation-based estimator of $V(\theta_0)$, let B denote a positive integer (e.g. $B = 500$) and suppose that $\hat{\theta}^{CL}$ is the maximum composite likelihood estimate of θ from (3.4) when applied to the original data. To estimate $V(\theta_0)$, do as follows:

1. For $b = 1, 2, \ldots, B$, simulate N observations of a trawl process $X^{(b)} = \{X_i^{(b)}\}_{i=1}^N$ with underlying parameters $\hat{\theta}^{CL}$.

2. For $b = 1, 2, \ldots, B$, use the simulated data $X^{(b)}$ to calculate $s^{(b)}(\hat{\theta}^{CL}) = N^{-1/2} \frac{\partial}{\partial \theta} l_{CL}(\hat{\theta}^{CL}; X^{(b)})$. The gradient can either be calculated numerically or analytically.[7] Note that the bootstrap data is used to calculate the gradient, but the parameter vector $\hat{\theta}^{CL}$ is the original estimator obtained from the initial (real) data set.

3. Estimate $V(\hat{\theta}^{CL})$ as the sample covariance matrix of the simulated scores $\left\{ s^{(b)}(\hat{\theta}^{CL}) \right\}_{b=1}^{B}$.

Note that the number of simulated observations, N, in Step 1 does not need to equal the number of observations in the original data set, n. When n is large, setting $N = n$ can be computationally costly; we found that setting $N = 500$ or even $N = 100$ provided good results. In our simulation study and in the empirical application we have set $B = N = 500$.

C Alternative expressions of the pairwise likelihood

Proposition 3.1 presented a simple expression for the pairwise PMFs $f(x_{i+k}, x_i; \theta)$ in the case where the underlying Lévy basis is positive. In the general case, i.e. where the Lévy basis is integer-valued, we have, by the law of total probability, that

$$\begin{aligned} f(x_{i+k},x_i;\theta) &:= \mathbb{P}_\theta \left(X_{(i+k)\Delta} = x_{i+k}, X_{i\Delta} = x_i \right) \\ &= \sum_{c=-\infty}^{\infty} \mathbb{P}_\theta \left(X_{(i+k)\Delta} = x_{i+k}, X_{i\Delta} = x_i | L(A_{(i+k)\Delta} \cap A_{i\Delta}) = c \right) \cdot \mathbb{P}_\theta \left(L(A_{(i+k)\Delta} \cap A_{i\Delta}) = c \right) \\ &= \sum_{c=-\infty}^{\infty} \mathbb{P}_\theta \left(L(A_{(i+k)\Delta} \setminus A_{i\Delta}) = x_{i+k} - c \right) \mathbb{P}_\theta \left(L(A_{i\Delta} \setminus A_{(i+k)\Delta}) = x_i - c \right) \\ &\quad \cdot \mathbb{P}_\theta \left(L(A_{(i+k)\Delta} \cap A_{i\Delta}) = c \right), \end{aligned} \tag{C.1}$$

which can be used for calculating the pairwise likelihood as a function of the parameter vector θ, in the general integer-valued case. When implementing this result in practice, one could truncate the sum in (C.1) according to some criterion in order to approximate the joint PMF. Truncation can be avoided by resorting to a simulation-based approach, however. The following proposition shows that a simulation unbiased version of the joint PMF exists, and that the simulation is, in fact, easy to perform.

Proposition C.1. *Let $t, s \geq 0$, choose $M \in \mathbb{N}$ and let $C^{(j)} \sim L(A_t \cap A_s)$, $j = 1, 2, \ldots, M$, be an iid sample. Then*

$$\hat{f}(x_t, x_s; \theta) = \frac{1}{M} \sum_{j=1}^{M} \mathbb{P}_\theta(L(A_t \setminus A_s) = x_t - C^{(j)}) \mathbb{P}_\theta(L(A_s \setminus A_t) = x_s - C^{(j)})$$

[7] The Supplementary Material contains analytical expressions for the gradients implied by the various parametric specifications considered in this paper.

is a simulation-based unbiased estimator of $f(x_t, x_s; \theta)$*. We further note that the simulation error*

$$\hat{f}(x_t, x_s; \theta) - f(x_t, x_s; \theta)$$

is, conditional on x*, stochastically independent for different values of* t *and* s*. Also, this error converges to zero at rate* $\sqrt{M}$ *as long as*

$$\sum_{c=-\infty}^{\infty} f(x_t|c;\theta)^2 f(x_s|c;\theta)^2 f(x;\theta) < \infty,$$

where $f(c;\theta) = \mathbb{P}_\theta(C^{(1)} = c)$ *denotes the PMF of* $C^{(j)}$*,* $j = 1, 2, \ldots, M$*, and* $f(x_t|c;\theta) = \mathbb{P}_\theta(X_t = x_t|L(A_t \cap A_s) = c)$ *denotes the conditional PMF of* X_t*.*

Proposition C.1 shows that the simulated CL function is an unbiased estimator of the true CL function. In other words, if we let U denote the vector of uniform random variables behind the simulation of $\{C^{(j)}\}_{j=1}^M$ and define

$$\log \mathcal{L}_U(\theta; x, u) = \log \mathcal{L}_U^{(K)}(\theta; x, u) := \sum_{k=1}^{K} \sum_{i=1}^{n-k} \log \hat{f}(x_{i+k}, x_i; \theta),$$

then $\mathcal{L}_U(\theta; x, u)$ is a simulation unbiased estimator for the composite likelihood $\mathcal{L}_{CL}(\theta; x)$. That is

$$\mathcal{L}_{CL}(\theta; x) = \int \mathcal{L}_U(\theta; x, u) f_U(u) du,$$

where $f_U(u) \propto 1$ is the joint density of the uniform random numbers behind all the simulation. It is well known that numerically optimizing a simulated likelihood function (the so-called simulated maximum likelihood approach, see e.g. Lerman and Manski, 1981) suffers a number of drawbacks and can be fragile in practice (e.g. Flury and Shephard, 2011). However, by a result in the seminal Andrieu et al. (2010), it is feasible to do Markov Chain Monte Carlo (MCMC) when one can unbiasedly simulate the likelihood. As a consequence, it is feasible to perform simulation-based estimation through MCMC, instead of relying on numerical optimization. From an estimation viewpoint, this can be an attractive approach (Flury and Shephard, 2011).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control 19*(6), 716–723.

Al-Osh, M. A. and A. A. Alzaid (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis 8*(3), 261–275.

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B 72*, 269–342.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: trades and quotes. *The Econometrics Journal 12*(3), C1–C32.

Barndorff-Nielsen, O. E., A. Lunde, N. Shephard, and A. Veraart (2014). Integer-valued trawl processes: A class of stationary infinitely divisible processes. *Scandinavian Journal of Statistics 41*, 693–724.

Barndorff-Nielsen, O. E., D. G. Pollard, and N. Shephard (2012). Integer-valued Lévy processes and low latency financial econometrics. *Quantitative Finance 12*, 587–605.

Bennedsen, M., A. Lunde, N. Shephard, and A. Veraart (2021). Inference and forecasting for continuous-time integer-valued trawl processes and their use in financial economics: Supplementary material. Preprint.

Bollen, N. P. B., T. Smith, and R. E. Whaley (2004). Modeling the bid/ask spread: measuring the inventory-holding premium. *Journal of Financial Economics 72*(1), 97–141.

Courgeau, V. and A. Veraart (2020). Asymptotic theory for the inference of the latent trawl model for extreme values. Available at SSRN: https://ssrn.com/abstract=3527739 or http://dx.doi.org/10.2139/ssrn.3527739.

Cox, D. R. and N. Reid (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika 91*(3), 729–737.

Davis, R., S. Holan, R. Lund, and N. Ravishanker (Eds.) (2016). *Handbook of Discrete-Valued Time Series*. Chapman and Hall/CRC.

Davis, R. A., T. do Rêgo Sousa, and C. Klüppelberg (2020). Indirect inference for time series using the empirical characteristic function and control variates.

Davis, R. A., K. Fokianos, S. H. Holan, H. Joe, J. Livsey, R. Lund, V. Pipiras, and N. Ravishanker (2021). Count time series: A methodological review. Forthcoming in *Journal of the American Statistical Association*.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Doukhan, P., K. Fokianos, and J. Rynkiewicz (2021). Mixtures of nonlinear Poisson autoregressions. *Journal of Time Series Analysis 42*(1), 107–135.

Doukhan, P., F. Roueff, and J. Rynkiewicz (2020). Spectral estimation for non-linear long range dependent discrete time trawl processes. *Electronic Journal of Statistics 14*(2), 3157–3191.

Elliott, G. and A. Timmermann (2016). *Economic Forecasting*. Princeton University Press.

Engle, R. F., C. Pakel, K. Sheppard, and N. Shephard (2020). Fitting vast dimensional time-varying covariance models. *Journal of Business and Economic Statistics*.

Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology and Climatology 8*(6), 985–987.

Flury, T. and N. Shephard (2011). Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. *Econometric Theory* (27), 933–956.

Fokianos, K. (2016). Statistical analysis of count time series models: a GLM perspective. In *Handbook of discrete-valued time series*, Chapman & Hall/CRC Handb. Mod. Stat. Methods, pp. 3–27. CRC Press, Boca Raton, FL.

Fokianos, K., A. Rahbek, and D. Tjøstheim (2009). Poisson autoregression. *Journal of the American Statistical Association 104*(488), 1430–1439.

Fokianos, K., B. Støve, D. Tjøstheim, and P. Doukhan (2020). Multivariate count autoregression. *Bernoulli 26*(1), 471–499.

Freeland, R. K. and B. P. M. McCabe (2004). Forecasting discrete valued low count time series. *International Journal of Forecasting 20*(3), 427–434.

Fuchs, F. and R. Stelzer (2013). Mixing conditions for multivariate infinitely divisible processes with an application to mixed moving averages and the supOU stochastic volatility model. *ESAIM: Probability and Statistics 17*, 455–471.

Gao, X. and P. X. K. Song (2010). Composite likelihood Bayesian information criteria for model selection in high-dimensional data. *Journal of the American Statistical Association 105*(492), 1531–1540.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood equation. *Ann. Math. Stat. 31*, 1208–1211.

Groß-KlußMann, A. and N. Hautsch (2013). Predicting bid–ask spreads using long-memory autoregressive conditional Poisson models. *Journal of Forecasting 32*(8), 724–742.

Hjort, N. L. and H. Omre (1994). Topics in spatial statistics (with discussion, comments and rejoinder). *Scandinavian Journal of Statistics 21*, 289–357.

Huang, R. D. and H. R. Stoll (1997). The components of the bid-ask spread: A general approach. *The Review of Financial Studies 10*(4), 995–1034.

Karlis, D. (2016). Models for multivariate count time series. In *Handbook of discrete-valued time series*, Chapman & Hall/CRC Handb. Mod. Stat. Methods, pp. 407–424. CRC Press, Boca Raton, FL.

Larribe, F. and P. Fearnhead (2011). On composite likelihoods in statistical genetics. *Statistica Sinica 21*(1), 43–69.

Lerman, S. and C. Manski (1981). On the use of simulated frequencies to approximate choice probabilities. In S. Lerman and C. Manski (Eds.), *Structural analysis of discrete data with econometric applications*, pp. 305–319. MIT Press.

Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics 80*, 220–239.

Livsey, J., R. Lund, S. Kechagias, and V. Pipiras (2018). Multivariate integer-valued time series with flexible autocovariances and their application to major hurricane counts. *Annals of Applied Statistics 12*(1), 408–431.

McCabe, B. and G. Martin (2005). Bayesian predictions of low count time series. *International Journal of Forecasting 21*(2), 315–330.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Journal of the American Water Resources Association 21*(4), 645–650.

Newey, W. K. and K. D. West (1987). a simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Ng, C. T. and H. Joe (2014). Model comparison with composite likelihood information criteria. *Bernoulli 20*(4), 1738–1764.

Noven, R., A. E. D. Veraart, and A. Gandy (2018). A latent trawl process model for extreme values. *Journal of Energy Markets 11*(3), 1–24.

Pakkanen, M. S., R. Passeggeri, O. Sauri, and A. E. D. Veraart (2021). Limit theorems for trawl processes. Forthcoming in *Electronic Journal of Probability*.

Pedeli, X., A. C. Davison, and K. Fokianos (2015). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *Journal of the American Statistical Association 110*(511), 1229–1238.

Pedeli, X. and D. Karlis (2013). Some properties of multivariate INAR(1) processes. *Computational Statistics & Data Analysis 67*, 213–225.

Schmidt, A. M. and J. B. M. Pereira (2011). Modelling time series of counts in epidemiology. *International Statistical Review 79*(1), 48–69.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464.

Shephard, N. and J. J. Yang (2016). Likelihood inference for exponential-trawl processes. In M. Podolskij, R. Stelzer, S. Thorbjørnsen, and A. E. D. Veraart (Eds.), *The Fascination of*

Probability, Statistics and their Applications: In Honour of Ole E. Barndorff-Nielsen, pp. 251–281. Springer International Publishing.

Shephard, N. and J. J. Yang (2017). Continuous time analysis of fleeting discrete price moves. *Journal of the American Statistical Association 112*(519), 1090–1106.

Silva, N., I. Pereira, and M. E. Silva (2009). Forecasting in INAR(1) model. *Revstat - Statistical Journal 7*(1), 119–134.

Takeuchi, K. (1976). Distribution of informational statistics and a criterion of model fitting. *Suri Kagaku [Mathematical Sciences] (in Japanese) 153*, 12–18.

Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis 92*(1), 1–28.

Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica 21*(1), 5–42.

Varin, C. and P. Vidoni (2005). A note on composite likelihood inference and model selection. *Biometrika 92*(3), 519–528.

Veraart, A. E. (2019). Modeling, simulation and inference for multivariate time series of counts using trawl processes. *Journal of Multivariate Analysis 169*, 110–129.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*(2), 307–333.

Wakefield, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics 8*(2), 158–183.

Supplemental Materials: Inference and forecasting for continuous-time integer-valued trawl processes and their use in financial economics

Mikkel Bennedsen[8], Asger Lunde[9], Neil Shephard[10], Almut E. D. Veraart[11]

Abstract

This document contains supplementary material for the article Bennedsen et al. (2021) including the proofs of the main theorems, additional simulation and empirical results, and more details on specific examples of trawl processes.

[8] Department of Economics and Business Economics and CREATES, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark. E-mail: mbennedsen@econ.au.dk

[9] Department of Economics and Business Economics and CREATES, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark. E-mail: alunde@econ.au.dk

[10] Department of Economics and Department of Statistics, Harvard University, One Oxford Street, Cambridge, MA 02138, USA. E-mail: shephard@fas.harvard.edu

[11] Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK and CREATES, Aarhus University. E-mail: a.veraart@imperial.ac.uk

1 Introduction

This document contains supplementary material for the article Bennedsen et al. (2021). It is structured as follows. Section 2 contains the proofs of our theoretical results. Section 3 contains additional simulation results for the MCL estimator: Section 3.1 presents simulation results supplementing those from the main paper, and Section 3.2 investigates the finite sample performance of the MCL estimator using a different simulation setup. Section 4 contains additional forecasting results, supplementing the results given in Section 6.1 of the main paper. Sections 5 and 6 present details on various parametric Lévy bases and trawl functions, respectively. These structures might be used in the construction of IVT processes, as illustrated in the main paper. Section 7 contains details on how to calculate the gradients for the log-composite likelihood functions implied by most of the parametric IVT processes. These calculations are straightforward to make (although somewhat tedious) and can rather easily be made for other IVT specifications than those considered here. The gradients can be used in the numerical optimization of the composite likelihood functions and are also crucial for implementing the asymptotic theory presented in the main paper. In particular, to estimate the asymptotic variance matrix $V(\theta)$, it is necessary to evaluate the gradient at $\hat{\theta}^{CL}$ (we refer to the Appendix of the main paper for details). Section 8 contains some additional technical calculations. Lastly, Section 9 contains brief details on the software packages accompanying the main paper. In particular, we supply software for simulation, estimation (including inference), model selection, and forecasting of IVT processes.

2 Mathematical proofs

We first give an alternative representation of the Lévy basis L, underlying the IVT process X. From the construction of the IVT process in Section 2 in Bennedsen et al. (2021), it is clear that the distribution of L is representable as a compound Poisson distribution. That is, for a Borel set B, we can write

$$\begin{aligned}\mathbb{P}_\theta(L(B)=x) &= \sum_{q=0}^{\infty}\mathbb{P}_\theta\left(\sum_{i=1}^{q}Y_i=x,\tilde{N}(B)=q\right)\\ &= \sum_{q=0}^{\infty}\mathbb{P}_\theta\left(\sum_{i=1}^{q}Y_i=x\right)\mathbb{P}_\theta(\tilde{N}(B)=q),\quad x\in\mathbb{Z},\end{aligned} \tag{S1}$$

where Y_i are iid integer-valued random variables with probability mass function $\tilde{\eta}(y) := \frac{\eta(y)}{\sum_{y=-\infty}^{\infty}\eta(y)}$, i.e. $\mathbb{P}_\theta(Y_i=y)=\tilde{\eta}(y)$, where η is the Lévy measure given in the construction of the IVT process in Equation (2.1) in Bennedsen et al. (2021). Likewise, $\tilde{N}$ is a Poisson random measure, given by

$$\tilde{N}(dx,ds)=\int_{-\infty}^{\infty}N(dy,dx,ds),$$

with an underlying intensity $\tilde{\nu} := \sum_{y=-\infty}^{\infty}\eta(y)$. The random variables Y_i are independent of the random measure $\tilde{N}$. Intuitively, we have decomposed the event that the sum of the points in the set B equals x (i.e. $\{L(B)=x\}$) into the intersection of the two events that there are q individual points in B (i.e. $\{\tilde{N}(B)=q\}$) and the "sizes" of these q points add up to x (i.e. $\{\sum_{i=1}^{q}Y_i=x\}$). With this construction, we have

$$\mathbb{P}_\theta(\tilde{N}(B)=q)=\frac{(\tilde{\nu}Leb(B))^q}{q!}e^{-\tilde{\nu}Leb(B)},\quad q=0,1,2,\ldots. \tag{S2}$$

We will use this alternative representation of L in our proofs below.

Proof of Theorem 3.1. Let $m>0$ and define $N_m=\tilde{N}(A_m\cap A)$ as the Poisson random variable, which counts the number of 'events' in the set $A_m\cap A_0$. From (S2), we know that there exists a constant $\tilde{\nu}>0$ such that

$$\mathbb{P}(N_m=0)=e^{-\tilde{\nu}Leb(A_m\cap A_0)}$$

and, therefore,

$$\mathbb{P}(N_m\neq 0)=1-\mathbb{P}(N_m=0)=1-e^{-\tilde{\nu}Leb(A_m\cap A_0)}=\tilde{\nu}Leb(A_m\cap A_0)+o(Leb(A_m\cap A_0)), \tag{S3}$$

as $m\to\infty$.

Let $G\in\mathcal{F}_{-\infty}^{0}$ and $H\in\mathcal{F}_{m}^{\infty}$ be such that $\mathbb{P}(G),\mathbb{P}(H)>0$, and write, using the law of total probability,

$$|\mathbb{P}(H\cap G)-\mathbb{P}(H)\mathbb{P}(G)|=|\mathbb{P}(H|G)-\mathbb{P}(H)|\cdot\mathbb{P}(G)$$

$$
\begin{aligned}
&= |\mathbb{P}(H|G, N_m = 0)\mathbb{P}(N_m = 0|G) + \mathbb{P}(H|G, N_m \neq 0)\mathbb{P}(N_m \neq 0|G) \\
&- \mathbb{P}(H|N_m = 0)\mathbb{P}(N_m = 0) - \mathbb{P}(H|N_m \neq 0)\mathbb{P}(N_m \neq 0)| \cdot \mathbb{P}(G) \\
&\leq (D_{1,m} + D_{2,m}) \cdot \mathbb{P}(G),
\end{aligned}
$$

where

$$
\begin{aligned}
D_{1,m} &:= |\mathbb{P}(H|G, N_m = 0)\mathbb{P}(N_m = 0|G) - \mathbb{P}(H|N_m = 0)\mathbb{P}(N_m = 0)|, \\
D_{2,m} &:= |\mathbb{P}(H|G, N_m \neq 0)\mathbb{P}(N_m \neq 0|G) - \mathbb{P}(H|N_m \neq 0)\mathbb{P}(N_m \neq 0)|.
\end{aligned}
$$

We seek to bound these expressions. For $D_{1,m}$, we use the fact that on the event $\{N_m = 0\}$ the two events G and H are independent. For both $D_{1,m}$ and $D_{2,m}$, we will use that the probability of the complementary event $\{N_m \neq 0\}$ is "small enough", cf. Equation (S3). For the first of the terms, we get, using conditional independence of H and G and (S3),

$$
\begin{aligned}
D_{1,m} &= \mathbb{P}(H|N_m = 0) \cdot |\mathbb{P}(N_m = 0|G) - \mathbb{P}(N_m = 0)| \\
&\leq |\mathbb{P}(N_m = 0|G) - \mathbb{P}(N_m = 0)| \\
&\leq |1 - \mathbb{P}(N_m = 0|G)| + |1 - \mathbb{P}(N_m = 0)| \\
&= |1 - \mathbb{P}(N_m = 0|G)| + \tilde{\nu} Leb(A_m \cap A_0) + o(Leb(A_m \cap A_0)),
\end{aligned}
$$

and, using Bayes formula and then the law of total probability,

$$
\begin{aligned}
|1 - \mathbb{P}(N_m = 0|G)| &= |1 - \mathbb{P}(G|N_m = 0)\mathbb{P}(N_m = 0)\mathbb{P}(G)^{-1}| \\
&= \mathbb{P}(G)^{-1}|\mathbb{P}(G) - \mathbb{P}(G|N_m = 0)\mathbb{P}(N_m = 0)| \\
&= \mathbb{P}(G)^{-1}|\mathbb{P}(G|N_m = 0)\mathbb{P}(N_m = 0) + \mathbb{P}(G|N_m \neq 0)\mathbb{P}(N_m \neq 0) \\
&- \mathbb{P}(G|N_m = 0)\mathbb{P}(N_m = 0)| \\
&= \mathbb{P}(G)^{-1}|\mathbb{P}(G|N_m \neq 0)\mathbb{P}(N_m \neq 0)| \\
&\leq \mathbb{P}(G)^{-1}\mathbb{P}(N_m \neq 0) \\
&= \mathbb{P}(G)^{-1}(\tilde{\nu} Leb(A_m \cap A_0) + o(Leb(A_m \cap A_0))).
\end{aligned}
$$

We conclude that

$$
D_{1,m} \cdot \mathbb{P}(G) \leq 2\tilde{\nu} Leb(A_m \cap A_0) + o(Leb(A_m \cap A_0)).
$$

For the second term above, we get

$$
\begin{aligned}
D_{2,m} &= |\mathbb{P}(H|G, N_m \neq 0)\mathbb{P}(N_m \neq 0|G) - \mathbb{P}(H|N_m \neq 0)\mathbb{P}(N_m \neq 0)| \\
&= |\mathbb{P}(H|G, N_m \neq 0)\mathbb{P}(N_m \neq 0|G) - \mathbb{P}(H|G, N_m \neq 0)\mathbb{P}(N_m \neq 0) \\
&+ \mathbb{P}(H|G, N_m \neq 0)\mathbb{P}(N_m \neq 0) - \mathbb{P}(H|N_m \neq 0)\mathbb{P}(N_m \neq 0)| \\
&\leq E_{1,m} + E_{2,m},
\end{aligned}
$$

where

$$E_{1,m} := |\mathbb{P}(H|G, N_m \neq 0)\mathbb{P}(N_m \neq 0|G) - \mathbb{P}(H|G, N_m \neq 0)\mathbb{P}(N_m \neq 0)|,$$
$$E_{2,m} := \ |\mathbb{P}(H|G, N_m \neq 0)\mathbb{P}(N_m \neq 0) - \mathbb{P}(H|N_m \neq 0)\mathbb{P}(N_m \neq 0)|.$$

Now, by (S3),

$$\begin{aligned} E_{2,m} &= \mathbb{P}(N_m \neq 0)\ |\mathbb{P}(H|G, N_m \neq 0) - \mathbb{P}(H|N_m \neq 0)| \\ &\leq \mathbb{P}(N_m \neq 0) \\ &= \tilde{\nu} Leb(A_m \cap A_0) + o(Leb(A_m \cap A_0)). \end{aligned}$$

Also,

$$\begin{aligned} E_{1,m} &= |\mathbb{P}(H|G, N_m \neq 0)\mathbb{P}(N_m \neq 0|G) - \mathbb{P}(H|G, N_m \neq 0)\mathbb{P}(N_m \neq 0)| \\ &= \mathbb{P}(H|G, N_m \neq 0) \cdot |\mathbb{P}(N_m \neq 0|G) - \mathbb{P}(N_m \neq 0)| \\ &\leq |\mathbb{P}(N_m \neq 0|G) - \mathbb{P}(N_m \neq 0)| \\ &\leq \mathbb{P}(N_m \neq 0|G) + \mathbb{P}(N_m \neq 0). \end{aligned}$$

Using Bayes formula, we can write

$$\begin{aligned} \mathbb{P}(N_m \neq 0|G) &= \mathbb{P}(G|N_m \neq 0)\mathbb{P}(N_m \neq 0)\mathbb{P}(G)^{-1} \\ &\leq \mathbb{P}(N_m \neq 0)\mathbb{P}(G)^{-1} \end{aligned}$$

so that, from (S3),

$$E_{1,m} \leq (1 + \mathbb{P}(G)^{-1})(\tilde{\nu} Leb(A_m \cap A_0) + o(Leb(A_m \cap A_0))).$$

We conclude that

$$D_{2,m} \cdot \mathbb{P}(G) \leq 3\tilde{\nu} Leb(A_m \cap A_0) + o(Leb(A_m \cap A_0)).$$

Taking it all together, we have that

$$\begin{aligned} |\mathbb{P}(H \cap G) - \mathbb{P}(H)\mathbb{P}(G)| &\leq (D_{1,m} + D_{2,m}) \cdot \mathbb{P}(G) \\ &\leq 5\tilde{\nu} Leb(A_m \cap A_0) + o(Leb(A_m \cap A_0)), \end{aligned}$$

implying, since G and H were arbitrary, that (taking supremums)

$$\alpha_m \leq 5\tilde{\nu} Leb(A_m \cap A_0) + o(Leb(A_m \cap A_0)).$$

To show that $\alpha_m = O(Leb(A_m \cap A_0))$, let $x_1, x_2 \in \mathbb{Z}$ and define the events $H := \{X_0 = x_1\}$ and $G := \{X_m = x_2\}$. We can show that

$$|\mathbb{P}(H \cap G) - \mathbb{P}(H)\mathbb{P}(G)| = O(Leb(A_m \cap A_0)), \qquad m \to \infty. \tag{S4}$$

Since, clearly, $H \in \mathcal{F}^0_{-\infty}$ and $G \in \mathcal{F}^\infty_m$, we conclude that $\alpha_m = O(Leb(A_m \cap A_0))$ as $m \to \infty$. Using Equation (2.10), i.e. the fact that $\rho(m) = Leb(A)^{-1} Leb(A_m \cap A_0)$, the proof is completed.[12]

□

Proof of Proposition 3.1. When $\nu(y) = 0$ for $y < 0$ we have $P_\theta(L(A_{(i+k)\Delta} \cap A_{i\Delta}) = c) = 0$ for $c < 0$. Further, since the maximal amount of events in $L(A_{(i+k)\Delta} \cap A_{i\Delta})$ is bounded by the number of events in x_{t+k} and x_t (no negative values in the trawl sets), we also have $P_\theta(L(A_{(i+k)\Delta} \cap A_{i\Delta}) = c) = 0$ for $c > \min\{x_{t+k}, x_t\}$. This, together with the discussion of the decomposition of trawl sets in Section 3.1.1 (cf. Figure 2) in Bennedsen et al. (2021), and the law of total probability, implies that

$$\begin{aligned} f(x_{i+k},x_i;\theta) &:= \mathbb{P}_\theta\left(X_{(i+k)\Delta} = x_{i+k}, X_{i\Delta} = x_i\right) \\ &= \sum_{c=-\infty}^{\infty} \mathbb{P}_\theta\left(X_{(i+k)\Delta} = x_{i+k}, X_{i\Delta} = x_i | L(A_{(i+k)\Delta} \cap A_{i\Delta}) = c\right) \cdot \mathbb{P}_\theta\left(L(A_{(i+k)\Delta} \cap A_{i\Delta}) = c\right) \\ &= \sum_{c=0}^{\max\{x_i, x_{i+k}\}} \mathbb{P}_\theta\left(L(A_{(i+k)\Delta} \setminus A_{i\Delta}) = x_{i+k} - c\right) \mathbb{P}_\theta\left(L(A_{i\Delta} \setminus A_{(i+k)\Delta}) = x_i - c\right) \\ &\quad \cdot \mathbb{P}_\theta\left(L(A_{(i+k)\Delta} \cap A_{i\Delta}) = c\right), \end{aligned}$$

as we wanted to show. □

Proof of Theorem 3.2. Due to the stationarity and ergodicity of the IVT processes considered in this paper, the normalized log-composite likelihood function will converge in probability to its population counterpart, i.e.

$$Q_n(\theta) := \frac{1}{n} l_{CL}(\theta; X) = \frac{1}{n} \sum_{k=1}^{K} \sum_{i=1}^{n-k} \log f(X_{(i+k)\Delta}, X_{i\Delta}; \theta) \xrightarrow{\mathbb{P}} \mathbb{E}\left[\sum_{k=1}^{K} \log f(X_{k\Delta}, X_0; \theta)\right] =: Q(\theta),$$

as $n \to \infty$. Since the pairwise likelihoods are indeed proper (bivariate) likelihoods, the information inequality implies that $Q(\theta)$ is uniquely maximized at $\theta = \theta_0$ (see, e.g., Lemma 2.2 in Newey and McFadden, 1994, p. 2124). The result now follows from Theorem 4.1 and Theorem 4.3 in Wooldridge (1994). □

Proof of Theorem 3.3. Let

$$s_n(\theta) := \frac{\partial}{\partial \theta} l_{CL}(\theta; X) = \sum_{k=1}^{K} \sum_{i=1}^{n-k} \frac{\partial}{\partial \theta} \log f(X_{(i+k)\Delta}, X_{i\Delta}; \theta)$$

denote the score function and consider the estimating equation related to the MCL estimator $\hat{\theta} = \hat{\theta}^{CL}$, namely $s_n(\hat{\theta}) = 0$. Using this equation, we Taylor expand $s_n(\hat{\theta})$ around the true parameter

[12] The result in Equation (S4) can be shown using similar arguments as those Lemma 2.1 below. Since (S4) is simpler to prove than Lemma 2.1, we omit the proof of the former. In fact, Lemma 2.1 implies that (S4) holds for the alternative sets $H' := \{X_0 = x_1, X_{-1} = x_2\}$ and $G' := \{X_m = x_3, X_{m+1} = x_4\}$. Since, also in this case, $H' \in \mathcal{F}^0_{-\infty}$ and $G' \in \mathcal{F}^\infty_m$, we see that Lemma 2.1 can indeed be directly applied to achieve the required result.

vector θ_0 to get

$$s_n(\theta_0) + \frac{\partial}{\partial \theta'} s_n(\bar{\theta})(\hat{\theta} - \theta_0) = 0,$$

where $\bar{\theta}$ lies on the line segment between θ_0 and $\hat{\theta}$ and $\frac{\partial}{\partial \theta'} s_n(\bar{\theta})$ is shorthand for $\frac{\partial}{\partial \theta'} s_n(\theta)|_{\theta=\bar{\theta}}$. Rearranging this equation and multiplying through by $\sqrt{n}$, we get

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\left(\frac{1}{n}\frac{\partial}{\partial \theta'} s_n(\bar{\theta})\right)^{-1} n^{-1/2} s_n(\theta_0).$$

Stationarity and ergodicity, along with consistency of $\hat{\theta}$ due to Theorem 3.2, implies that $-\frac{1}{n}\frac{\partial}{\partial \theta'} s_n(\bar{\theta})' \xrightarrow{\mathbb{P}} H(\theta_0)$ as $n \to \infty$. To prove the result, we thus need to show that $n^{-1/2} s_n(\theta_0) \xrightarrow{(d)} N(0, V(\theta_0))$ as $n \to \infty$. By the mixing properties of the IVT process X, given in Theorem 3.1, it is enough to show that $\mathbb{E}[s_n(\theta_0)] = 0$ and $Var\left(n^{-1/2} s_n(\theta_0)\right) \to V(\theta_0)$ as $n \to \infty$ (e.g. Davidson, 1994, Corollary 24.7, p. 387).[13]

To show this, we consider, for simplicity, the case where θ is a scalar. The vector case is similar, but with slightly more involved notation. First note that, clearly,

$$\mathbb{E}\left[s_n(\theta_0)\right] = \mathbb{E}\left[\frac{\partial}{\partial \theta} l_{CL}(\theta; X)|_{\theta=\theta_0}\right] = 0.$$

Also

$$\begin{aligned}
Var\left(s_n(\theta_0)\right) &= Var\left(\frac{\partial}{\partial \theta} l_{CL}(\theta; X)|_{\theta=\theta_0}\right) \\
&= Var\left(\sum_{k=1}^{K}\sum_{i=1}^{n-k} \frac{\partial}{\partial \theta} \log f(X_{i\Delta}, X_{(i+k)\Delta}; \theta)\big|_{\theta=\theta_0}\right) \\
&= \sum_{k=1}^{K}\sum_{i=1}^{n-k} Var\left(\frac{\partial}{\partial \theta} \log f(X_{i\Delta}, X_{(i+k)\Delta}; \theta)\big|_{\theta=\theta_0}\right) \\
&+ \sum_{k=1}^{K}\sum_{k'=1}^{K}\sum_{i\neq j} Cov\left(\frac{\partial}{\partial \theta} \log f(X_{i\Delta}, X_{(i+k)\Delta}; \theta)\big|_{\theta=\theta_0}, \frac{\partial}{\partial \theta} \log f(X_{j\Delta}, X_{(j+k')\Delta}; \theta)\big|_{\theta=\theta_0}\right).
\end{aligned}$$

Due to stationarity, the first sum is $O(n)$ as $n \to \infty$. To prove the proposition, we therefore investigate the second sum. With slight abuse of notation, let $\frac{\partial}{\partial \theta} f(X_{i\Delta}, X_{(i+k)\Delta}; \theta)\big|_{\theta=\theta_0}$ be denoted by $\frac{\partial}{\partial \theta} \log f(X_{i\Delta}, X_{(i+k)\Delta}; \theta_0)$. For $l \geq 1$, define also the joint probability mass function

$$f_l(x_1, x_2, x_3, x_4; \theta, k, k') := \mathbb{P}_\theta\left(X_0 = x_1, X_{k\Delta} = x_2, X_{l\Delta} = x_3, X_{(l+k')\Delta} = x_4\right), \quad x_1, x_2, x_3, x_4 \in \mathbb{Z}.$$

Now, using that

$$\mathbb{E}\left[\frac{\partial}{\partial \theta} \log f(X_{i\Delta}, X_{(i+k)\Delta}; \theta_0)\right] = 0,$$

[13] Note that the crucial condition (c') of Corollary 24.7 in Davidson (1994) relies on the IVT process X being mixing of size ϕ_0 for some $\phi_0 > 1$. This rules out the long memory processes, as shown in Theorem 3.1.

we have, for all i, j, k, k',

$$
\begin{aligned}
&Cov\left(\frac{\partial}{\partial\theta}\log f(X_{i\Delta}, X_{(i+k)\Delta};\theta_0), \frac{\partial}{\partial\theta}\log f(X_{j\Delta}, X_{(j+k')\Delta};\theta_0)\right)\\
&= \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(X_{i\Delta}, X_{(i+k)\Delta};\theta_0)\frac{\partial}{\partial\theta}\log f(X_{j\Delta}, X_{(j+k')\Delta};\theta_0)\right]\\
&= \sum_{x_1=-\infty}^{\infty}\sum_{x_2=-\infty}^{\infty}\sum_{x_3=-\infty}^{\infty}\sum_{x_4=-\infty}^{\infty}\frac{\partial}{\partial\theta}f(x_1,x_2;\theta_0)\frac{\partial}{\partial\theta}f(x_3,x_4;\theta_0)\frac{f_{|i-j|}(x_1,x_2,x_3,x_4;\theta_0,k,k')}{f(x_1,x_2;\theta_0)f(x_3,x_4;\theta_0)}\\
&= \sum_{x_1=-\infty}^{\infty}\sum_{x_2=-\infty}^{\infty}\sum_{x_3=-\infty}^{\infty}\sum_{x_4=-\infty}^{\infty}\frac{\partial}{\partial\theta}f(x_1,x_2;\theta_0)\frac{\partial}{\partial\theta}f(x_3,x_4;\theta_0)\left(\frac{f_{|i-j|}(x_1,x_2,x_3,x_4;\theta_0,k,k')}{f(x_1,x_2;\theta_0)f(x_3,x_4;\theta_0)}-1\right),
\end{aligned}
$$

where the last equality follows because, e.g.,

$$
\sum_{x_1=-\infty}^{\infty}\sum_{x_2=-\infty}^{\infty}\frac{\partial}{\partial\theta}f(x_1,x_2;\theta_0) = \frac{\partial}{\partial\theta}\sum_{x_1=-\infty}^{\infty}\sum_{x_2=-\infty}^{\infty}f(x_1,x_2;\theta_0) = \frac{\partial}{\partial\theta}1 = 0.
$$

Now, Lemma 2.1 below shows that

$$
\left(\frac{f_n(x_1,x_2,x_3,x_4;\theta_0,k,k')}{f(x_1,x_2;\theta_0)f(x_3,x_4;\theta_0)}-1\right) = O(Leb(A_{n\Delta}\cap A_0)), \quad n\to\infty,
$$

from which we conclude, using Equation (2.10), i.e. $Leb(A_{n\Delta}\cap A_0) = \rho(n\Delta)Leb(A)$, and the summability condition on ρ imposed in the theorem, that the second sum in the expression for $Var\left(s_n(\theta_0)\right)$ is $O(n)$ as well. Indeed, taking it all together, we have that, as $n\to\infty$,

$$
\begin{aligned}
n^{-1}Var\left(s_n(\theta_0)\right) \to &\sum_{k=1}^{K}Var\left(\frac{\partial}{\partial\theta}\log f(X_0, X_{k\Delta};\theta)|_{\theta=\theta_0}\right)\\
&+2\sum_{k=1}^{K}\sum_{k'=1}^{K}\sum_{i=1}^{\infty}Cov\left(\frac{\partial}{\partial\theta}\log f(X_0, X_{k\Delta};\theta)|_{\theta=\theta_0}, \frac{\partial}{\partial\theta}\log f(X_{i\Delta}, X_{(i+k')\Delta};\theta)\big|_{\theta=\theta_0}\right)\\
&:= V(\theta_0),
\end{aligned}
$$

where the series converges. This finalizes the proof.

□

Proof of Theorem 3.4. With similar calculations to those used in the proof of Theorem 3.3, we can write

$$
n^{H/2}(\hat\theta-\theta_0) = -\left(\frac{1}{n}\frac{\partial}{\partial\theta'}s_n(\bar\theta)\right)^{-1}n^{H/2-1}s_n(\theta_0).
$$

We again have that $-\frac{1}{n}\frac{\partial}{\partial\theta}s_n(\bar\theta)' \xrightarrow{\mathbb{P}} H(\theta_0)$ as $n\to\infty$. As in the proof of Theorem 3.3, we can write

$$
Var\left(s_n(\theta_0)\right) = A_{1,n} + A_{2,n},
$$

where $A_{1,n}$ is $O(n)$. Further, Lemma 2.1 below implies that

$$
A_{2,n} = O(n^2 Leb(A_{n\Delta}\cap A_0)),
$$

as $n \to \infty$. Equation (2.10) along with the condition on ρ imposed in the theorem thus yields

$$A_{2,n} = O(n^2 L_\infty(n\Delta) n^{-H}),$$

as $n \to \infty$. Using this, we get that, for all $\epsilon > 0$,

$$n^{H-2\pm 2\epsilon} Var\left(s_n(\theta_0)\right) = O(n^{H-1\pm 2\epsilon}) + O(L_\infty(n\Delta) n^{\pm 2\epsilon}), \tag{S5}$$

as $n \to \infty$. Finally, we recall the so-called *Potter bounds* for slowly varying functions: Since L_∞ is a slowly varying function, for all $\epsilon > 0$ it holds that (Bingham et al., 1989, Theorem 1.5.6(ii))

$$L_\infty(n\Delta) n^{\epsilon} \to \infty \quad \text{and} \quad L_\infty(n\Delta) n^{-\epsilon} \to 0,$$

as $n \to \infty$. Combining the Potter bounds with (S5) yields the required results. □

Lemma 2.1. *Fix $k, k' \geq 1$, let X be an IVT process, let $f(\cdot,\cdot;\theta)$ be the joint PMF of $(X_0, X_{k\Delta})$, and let $f_n(\cdot,\cdot,\cdot,\cdot;\theta)$ be the joint PMF of $(X_0, X_{k\Delta}, X_{n\Delta}, X_{(n+k')\Delta})$. That is*

$$f(x_1, x_2; \theta) := \mathbb{P}_\theta\left(X_0 = x_1, X_{k\Delta} = x_2\right), \quad x_1, x_2 \in \mathbb{Z},$$

and

$$f_n(x_1, x_2, x_3, x_4; \theta) := \mathbb{P}_\theta\left(X_0 = x_1, X_{k\Delta} = x_2, X_{n\Delta} = x_3, X_{(n+k')\Delta} = x_4\right), \quad x_1, x_2, x_3, x_4 \in \mathbb{Z}.$$

Define the function

$$G_n(x_1, x_2, x_3, x_4; \theta) := \left(\frac{f_n(x_1, x_2, x_3, x_4; \theta)}{f(x_1, x_2; \theta) f(x_3, x_4; \theta)} - 1\right).$$

The following holds:

$$\frac{1}{Leb(A_{k\Delta} \cap A_{n\Delta})} G_n(x_1, x_2, x_3, x_4; \theta) \to G(x_1, x_2, x_3, x_4; \theta), \quad n \to \infty,$$

where G is a function, given in Equation (S9) below, that depends on the Lévy basis and trawl function of X. (In Remark 2.1 below, we give the function G in the special case where the Lévy basis is Poissonian and the trawl function is the Gamma trawl.)

Proof of Lemma 2.1. Letting $f_n(x_1, x_2 | x_3, x_4; \theta) := \mathbb{P}_\theta\left(X_0 = x_1, X_{k\Delta} = x_2 | X_{n\Delta} = x_3, X_{(n+k')\Delta} = x_4\right)$, we can write

$$\begin{aligned} G_n(x_1, x_2, x_3, x_4; \theta) &= \frac{f_n(x_1, x_2, x_3, x_4; \theta) - f(x_1, x_2; \theta) f(x_3, x_4; \theta)}{f(x_1, x_2; \theta) f(x_3, x_4; \theta)} \\ &= \frac{f_n(x_1, x_2 | x_3, x_4; \theta) f(x_3, x_4; \theta) - f(x_1, x_2; \theta) f(x_3, x_4; \theta)}{f(x_1, x_2; \theta) f(x_3, x_4; \theta)} \\ &= \left(f_n(x_1, x_2 | x_3, x_4; \theta) - f(x_1, x_2; \theta)\right) f(x_1, x_2; \theta)^{-1}. \end{aligned}$$

To prove the lemma, we therefore study the asymptotic behavior of $f_n(x_1, x_2 | x_3, x_4; \theta) - f(x_1, x_2; \theta)$ as $n \to \infty$.

Recall first the decomposition of the trawl sets into three disjoint sets which led to Proposition 3.1, see Figure 2, in Bennedsen et al. (2021). In a similar manner, we can decompose the four trawl sets associated to $X_0 = L(A_0)$, $X_{k\Delta} = L(A_{k\Delta})$, $X_{n\Delta} = L(A_{n\Delta})$, and $X_{(n+k')\Delta} = L(A_{(n+k')\Delta})$, into 10 disjoint sets as illustrated in Figure S1 below. For ease of notation, we ignore the dependence on n, k, and k' for a moment and write

$$A_0 = C_3 \cup C_4 \cup C_6 \cup C_7, \quad A_{k\Delta} = C_1 \cup C_2 \cup C_3 \cup C_4 \cup C_6,$$
$$A_{n\Delta} = C_1 \cup C_2 \cup C_3 \cup C_4 \cup D_2 \cup D_3, \quad A_{(n+k')\Delta} = C_1 \cup C_3 \cup D_1 \cup D_2,$$

where the sets $C_1, C_2, \ldots, C_7, D_1, D_2, D_3$ are disjoint. We will use below that $\lim_{n\to\infty} Leb(C_j) = 0$ for $j = 1, 2, 3, 4$, $\lim_{n\to\infty} Leb(C_5) = Leb(A_{k\Delta} \setminus A_0)$, $\lim_{n\to\infty} Leb(C_6) = Leb(A_0 \cap A_{k\Delta})$, $\lim_{n\to\infty} Leb(C_7) = Leb(A_0 \setminus A_{k\Delta})$, $\lim_{n\to\infty} Leb(D_1) = Leb(A_{(n+k')\Delta} \setminus A_{n\Delta})$, $\lim_{n\to\infty} Leb(D_2) = Leb(A_{n\Delta} \cap A_{(n+k')\Delta})$, and $\lim_{n\to\infty} Leb(D_3) = Leb(A_{n\Delta} \setminus A_{(n+k')\Delta})$, cf. Figure S1 in Bennedsen et al. (2021).

Using this decomposition and the law of total probability, we may write

$$f(x_1, x_2; \theta) = \sum_{c_1,c_2,c_3,c_4=-\infty}^{\infty} \mathbb{P}_\theta \left(L(C_6) + L(C_7) = x_1 - c_3 - c_4, L(C_5) + L(C_6) = x_2 - c_1 - c_2 - c_3 - c_4\right) \cdot \prod_{j=1}^{4} \mathbb{P}_\theta \left(L(C_j) = c_j\right)$$

and

$$f(x_1, x_2 | x_3, x_4; \theta) = \sum_{c_1,c_2,c_3,c_4=-\infty}^{\infty} \mathbb{P}_\theta \left(L(C_6) + L(C_7) = x_1 - c_3 - c_4, L(C_5) + L(C_6) = x_2 - c_1 - c_2 - c_3 - c_4\right) \cdot \mathbb{P}_\theta \left(L(C_1) = c_1, L(C_2) = c_2, L(C_3) = c_3, L(C_4) = c_4 | X_{n\Delta} = x_3, X_{(n+k')\Delta} = x_4\right).$$

Taking these together, we get

$$\begin{aligned} &f_n(x_1, x_2 | x_3, x_4; \theta) - f(x_1, x_2; \theta) \\ &= \sum_{c_1,c_2,c_3,c_4=-\infty}^{\infty} \mathbb{P}_\theta \left(L(C_6) + L(C_7) = x_1 - c_3 - c_4, L(C_5) + L(C_6) = x_2 - c_1 - c_2 - c_3 - c_4\right) \\ &\cdot \left(\mathbb{P}_\theta \left(L(C_1) = c_1, L(C_2) = c_2, L(C_3) = c_3, L(C_4) = c_4 | X_{n\Delta} = x_3, X_{(n+k')\Delta} = x_4\right) - \prod_{j=1}^{4} \mathbb{P}_\theta \left(L(C_j) = c_j\right) \right). \end{aligned}$$

Note that, for the first term in the parenthesis, the following holds

$$\begin{aligned} &\mathbb{P}_\theta \left(L(C_1) = c_1, L(C_2) = c_2, L(C_3) = c_3, L(C_4) = c_4 | X_{n\Delta} = x_3, X_{(n+k')\Delta} = x_4\right) \\ &= f(x_3, x_4; \theta)^{-1} \mathbb{P}_\theta \left(L(C_1) = c_1, L(C_2) = c_2, L(C_3) = c_3, L(C_4) = c_4, X_{n\Delta} = x_3, X_{(n+k')\Delta} = x_4\right) \\ &= f(x_3, x_4; \theta)^{-1} \mathbb{P}_\theta \left(L(D_2) + L(D_3) = x_3 - c_1 - c_2 - c_3 - c_4, L(D_1) + L(D_2) = x_4 - c_1 - c_3\right) \\ &\cdot \prod_{j=1}^{4} \mathbb{P}_\theta \left(L(C_j) = c_j\right), \end{aligned}$$

which allows us to write

$$
\begin{aligned}
&f_n(x_1,x_2|x_3,x_4;\theta) - f(x_1,x_2;\theta) \\
&= \sum_{c_1,c_2,c_3,c_4=-\infty}^{\infty} \mathbb{P}_\theta\left(L(C_6)+L(C_7)=x_1-c_3-c_4, L(C_5)+L(C_6)=x_2-c_1-c_2-c_3-c_4\right) \\
&\cdot \prod_{j=1}^{4} \mathbb{P}_\theta\left(L(C_j)=c_j\right) f(x_3,x_4;\theta)^{-1} \\
&\cdot \left(\mathbb{P}_\theta\left(L(D_2)+L(D_3)=x_3-c_1-c_2-c_3-c_4, L(D_1)+L(D_2)=x_4-c_1-c_3\right) - f(x_3,x_4)\right).
\end{aligned}
$$

Define the set $\mathcal{C}_0 := \{(c_1,c_2,c_3,c_4) \in \mathbb{Z}^4 : c_i \neq 0 \text{ for at least one } i = 1,2,3,4\}$. The above calculations imply that

$$
f_n(x_1,x_2|x_3,x_4;\theta) - f(x_1,x_2;\theta) = f(x_3,x_4)^{-1}\left(F_n^{(1)} + F_n^{(2)}\right), \tag{S6}
$$

where

$$
\begin{aligned}
F_n^{(1)} &:= \mathbb{P}_\theta\left(L(C_6)+L(C_7)=x_1, L(C_5)+L(C_6)=x_2\right) \prod_{j=1}^{4} \mathbb{P}_\theta\left(L(C_j)=0\right) \\
&\quad \cdot \left(\mathbb{P}_\theta\left(L(D_2)+L(D_3)=x_3, L(D_1)+L(D_2)=x_4\right) - f(x_3,x_4;\theta)\right)
\end{aligned}
$$

and

$$
\begin{aligned}
F_n^{(2)} &:= \sum_{(c_1,c_2,c_3,c_4)\in\mathcal{C}_0} \mathbb{P}_\theta\left(L(C_6)+L(C_7)=x_1-c_3-c_4, L(C_5)+L(C_6)=x_2-c_1-c_2-c_3-c_4\right) \\
&\quad \cdot \prod_{j=1}^{4} \mathbb{P}_\theta\left(L(C_j)=c_j\right) \\
&\quad \cdot \left(\mathbb{P}_\theta\left(L(D_2)+L(D_3)=x_3-c_1-c_2-c_3-c_4, L(D_1)+L(D_2)=x_4-c_1-c_3\right) - f(x_3,x_4;\theta)\right).
\end{aligned}
$$

We can think of $F_n^{(1)}$ as the part of $(f_n(x_1,x_2|x_3,x_4;\theta) - f(x_1,x_2;\theta))f(x_3,x_4)$ where $c_1 = c_2 = c_3 = c_4 = 0$, while $F_n^{(2)}$ is the remainder.

We study first the behavior of $F_n^{(1)}$ as $n \to \infty$. Considering the first two factors of this term, the continuity of the probability measure $\mathbb{P}_\theta(\cdot)$ implies that

$$
\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}_\theta\left(L(C_6)+L(C_7)=x_1, L(C_5)+L(C_6)=x_2\right) \prod_{j=1}^{4} \mathbb{P}_\theta\left(L(C_j)=0\right) \\
&= \mathbb{P}_\theta\left(L(A_0\cap A_{k\Delta})+L(A_0\setminus A_{k\Delta})=x_1, L(A_0\cap A_{k\Delta})+L(A_{k\Delta}\setminus A_0)=x_2\right) \\
&= f(x_1,x_2;\theta).
\end{aligned}
$$

The third term in $F_n^{(1)}$, i.e.

$$
\mathbb{P}_\theta\left(L(D_2)+L(D_3)=x_3, L(D_1)+L(D_2)=x_4\right) - f(x_3,x_4;\theta)
$$

will, by the same logic as above, converge to zero as $n \to \infty$. In fact, by decomposition of the trawl sets of $f(x_3, x_4; \theta)$ in the same manner as above, we get that

$$\begin{aligned}
&\mathbb{P}_\theta \left(L(D_2) + L(D_3) = x_3, L(D_1) + L(D_2) = x_4 \right) - f(x_3, x_4; \theta) \\
&= \mathbb{P}_\theta \left(L(D_2) + L(D_3) = x_3, L(D_1) + L(D_2) = x_4 \right) \left(1 - \prod_{j=1}^{4} \mathbb{P}_\theta \left(L(C_j) = 0 \right) \right) \\
&- \sum_{(c_1,c_2,c_3,c_4) \in \mathcal{C}_0} \mathbb{P}_\theta \left(L(D_2) + L(D_3) = x_3 - c_1 - c_2 - c_3 - c_4, L(D_1) + L(D_2) = x_4 - c_1 - c_3 \right) \\
&\cdot \prod_{j=1}^{4} \mathbb{P}_\theta \left(L(C_j) = c_j \right).
\end{aligned}$$

In the first part of Lemma 2.2 below, we show that there exists a constant $\tilde{\nu} > 0$, such that

$$1 - \prod_{j=1}^{4} \mathbb{P}_\theta \left(L(C_j) = 0 \right) = \tilde{\nu} Leb(A_{k\Delta} \cap A_{n\Delta}) + o\left(Leb(A_{k\Delta} \cap A_{n\Delta}) \right),$$

as $n \to \infty$. Similarly, in the second part of Lemma 2.2, we show that there exists a non-negative function $\tilde{\eta}$ concentrated on the integers, such that, for $c_j \neq 0$,

$$\mathbb{P}_\theta \left(L(C_j) = c_j \right) = \tilde{\eta}(c_j) \tilde{\nu} Leb(C_j) e^{-\tilde{\nu} Leb(C_j)} + o\left(Leb(C_j) \right),$$

as $n \to \infty$, while

$$\mathbb{P}_\theta \left(L(C_j) = 0 \right) = e^{-\tilde{\nu} Leb(C_j)} + o\left(Leb(C_j) \right),$$

as $n \to \infty$. This shows that for quadruplets (c_1, c_2, c_3, c_4) where $c_i \neq 0$ for some $i = 1, 2, 3, 4$ and $c_j = 0$ for the remaining $j \neq i$ (i.e. quadruplets of the form $(c_1, 0, 0, 0)$, $(0, c_2, 0, 0)$, $(0, 0, c_3, 0)$ or $(0, 0, 0, c_4)$), we have

$$\begin{aligned}
\prod_{j=1}^{4} \mathbb{P}_\theta \left(L(C_j) = c_j \right) &= \tilde{\eta}(c_i) \tilde{\nu} Leb(C_i) e^{-\tilde{\nu} \sum_{j=1}^{4} Leb(C_j)} + o\left(\sum_{j=1}^{4} Leb(C_j) \right) \\
&= \tilde{\eta}(c_i) \tilde{\nu} Leb(C_i) e^{-\tilde{\nu} Leb(A_{k\Delta} \cap A_{n\Delta})} + o\left(Leb(A_{k\Delta} \cap A_{n\Delta}) \right),
\end{aligned}$$

as $n \to \infty$. Conversely, for quadruplets (c_1, c_2, c_3, c_4) where $c_i, c_j \neq 0$ for at least two distinct $i, j = 1, 2, 3, 4$, we have

$$\prod_{j=1}^{4} \mathbb{P}_\theta \left(L(C_j) = c_j \right) = o\left(Leb(A_{k\Delta} \cap A_{n\Delta}) \right),$$

as $n \to \infty$.

Define the numbers $a_j := \lim_{n\to\infty} \frac{Leb(C_j)}{Leb(A_{n\Delta} \cap A_{k\Delta})} \geq 0$, $j = 1, 2, 3, 4$, which are such that $\sum_{j=1}^{4} a_j = 1$ since $C_1 \cup C_2 \cup C_3 \cup C_4 = A_{n\Delta} \cap A_{k\Delta}$, cf. Figure S1 below. Taking the above together, we may, after a little algebra, conclude that, as $n \to \infty$,

$$\frac{F_n^{(1)}}{Leb(A_{k\Delta} \cap A_{n\Delta})} \to \tilde{\nu} f(x_1, x_2; \theta) f(x_3, x_4; \theta) \tag{S7}$$

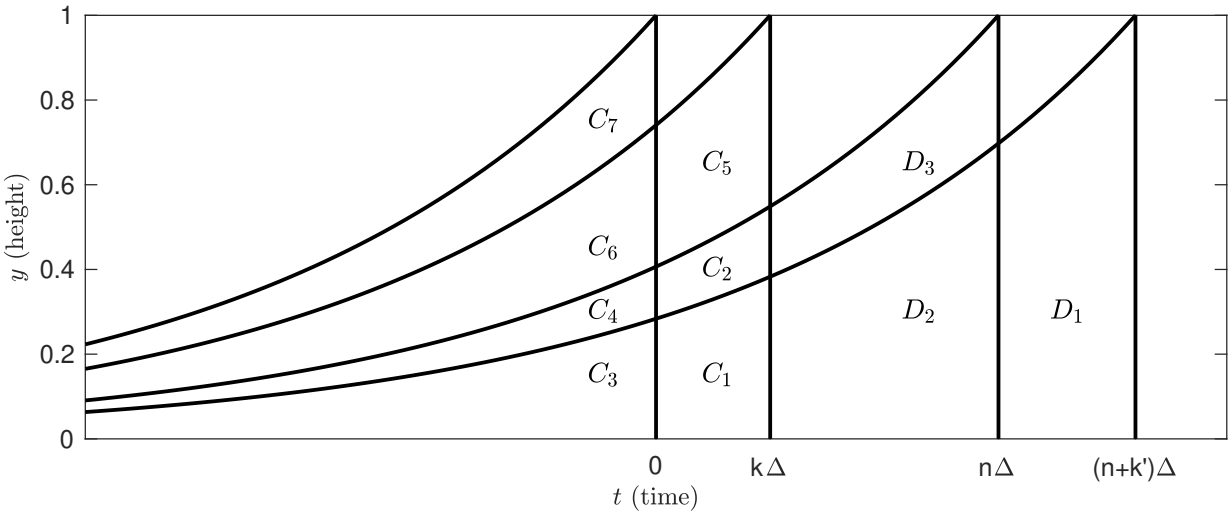

Figure S1: *Illustration of the decomposition of the four trawl sets* $A_0, A_{k\Delta}, A_{n\Delta}, A_{(n+k')\Delta}$.

$$- \tilde{\nu} f(x_1, x_2) \sum_{c \neq 0} \tilde{\eta}(c) \left((a_1 + a_3) f(x_3 - c, x_4 - c; \theta) + (a_2 + a_4) f(x_3 - c, x_4; \theta) \right).$$

Turning now to $F_n^{(2)}$, similar calculations yield that, as $n \to \infty$,

$$\begin{aligned}
\frac{F_n^{(2)}}{Leb(A_{k\Delta} \cap A_{n\Delta})} \to & - \tilde{\nu} f(x_3, x_4; \theta) \sum_{c \neq 0} \tilde{\eta}(c) \left((a_1 + a_2) f(x_1, x_2 - c; \theta) + (a_3 + a_4) f(x_1 - c, x_2 - c; \theta) \right) \\
& + \tilde{\nu} a_1 \sum_{c \neq 0} \tilde{\eta}(c) f(x_1, x_2 - c) f(x_3 - c, x_4 - c) \\
& + \tilde{\nu} a_2 \sum_{c \neq 0} \tilde{\eta}(c) f(x_1, x_2 - c) f(x_3 - c, x_4) \\
& + \tilde{\nu} a_3 \sum_{c \neq 0} \tilde{\eta}(c) f(x_1 - c, x_2 - c) f(x_3 - c, x_4 - c) \\
& + \tilde{\nu} a_4 \sum_{c \neq 0} \tilde{\eta}(c) f(x_1 - c, x_2 - c) f(x_3 - c, x_4). \qquad \text{(S8)}
\end{aligned}$$

Finally, recalling Equation (S6), we can conclude that

$$\begin{aligned}
\lim_{n \to \infty} \frac{G_n(x_1, x_2, x_3, x_4; \theta)}{Leb(A_{k\Delta} \cap A_{n\Delta})} &= \lim_{n \to \infty} \frac{F_n^{(1)} + F_n^{(2)}}{Leb(A_{k\Delta} \cap A_{n\Delta})} f(x_1, x_2; \theta)^{-1} f(x_3, x_4; \theta)^{-1} \\
&=: G(x_1, x_2, x_3, x_4; \theta), \qquad \text{(S9)}
\end{aligned}$$

where $\lim_{n \to \infty} \frac{1}{Leb(A_{k\Delta} \cap A_{n\Delta})} F_n^{(i)}$ for $i = 1, 2$ are given above in Equations (S7)–(S8). (See the following Remark 2.1 for how the expression for G simplifies slightly in the case of an IVT process with Poisson Lévy basis and Gamma trawl function.)

□

Remark 2.1. Note that in the case of a Poisson Lévy basis (Example 2.1 in Bennedsen et al. (2021)), we have $\tilde{\nu} = \nu$, $\tilde{\eta}(1) = 1$, and $\tilde{\eta}(c) = 0$ for $c \neq 1$. Further, in the case of d being a Gamma trawl function (Example 2.5), it is straightforward to show that $a_1 = a_2 = a_4 = 0$ and $a_3 = 1$. For this specification, the limit in the proof of Lemma 2.1 simplifies somewhat. Indeed, in this case, Equation (S9) yields

$$G(x_1, x_2, x_3, x_4; \theta) = \nu \left(\frac{f(x_1 - 1, x_2 - 1; \theta)}{f(x_1, x_2; \theta)} - 1 \right) \left(\frac{f(x_3 - 1, x_4 - 1; \theta)}{f(x_3, x_4; \theta)} - 1 \right).$$

Lemma 2.2. *In the setting of the proof of Lemma 2.1, we have that the following two-part result.*

(First part) There exists a constant $\tilde{\nu} > 0$, such that

$$1 - \prod_{j=1}^{4} \mathbb{P}_\theta \left(L(C_j) = 0 \right) = \tilde{\nu} Leb(A_{k\Delta} \cap A_{n\Delta}) + o\left(Leb(A_{k\Delta} \cap A_{n\Delta}) \right),$$

as $n \to \infty$.

(Second part) There exists a non-negative function $\tilde{\eta}$, concentrated on the integers, such that, for $c_j \neq 0$,

$$\mathbb{P}_\theta \left(L(C_j) = c_j \right) = \tilde{\eta}(c_j) \tilde{\nu} Leb(C_j) e^{-\tilde{\nu} Leb(C_j)} + o\left(Leb(C_j) \right),$$

as $n \to \infty$, while

$$\mathbb{P}_\theta \left(L(C_j) = 0 \right) = e^{-\tilde{\nu} Leb(C_j)} + o\left(Leb(C_j) \right),$$

as $n \to \infty$.

Proof of Lemma 2.2. The proof of the lemma relies on the alternative representation of the Lévy basis L given at the start of the appendix, see Equation (S1).

(Proof of second part) Note that since $\eta(0) = 0$, we have $\mathbb{P}_\theta \left(Y_i = 0 \right) = 0$. Using this in the setup of Lemma 2.1, we get, from Equations (S1) and (S2),

$$\begin{aligned} \mathbb{P}_\theta \left(L(C_j) = 0 \right) &= e^{-\tilde{\nu} Leb(C_j)} + \sum_{q=2}^{\infty} \mathbb{P}_\theta \left(\sum_{i=1}^{q} Y_i = x \right) \mathbb{P}_\theta(\tilde{N}(C_j) = q) \\ &= e^{-\tilde{\nu} Leb(C_j)} + o\left(Leb(C_j) \right), \end{aligned}$$

as $n \to \infty$, while for $c_j \neq 0$,

$$\begin{aligned} \mathbb{P}_\theta \left(L(C_j) = c_j \right) &= \mathbb{P}_\theta \left(Y_1 = c_j \right) \mathbb{P}_\theta(\tilde{N}(C_j) = 1) + \sum_{q=2}^{\infty} \mathbb{P}_\theta \left(\sum_{i=1}^{q} Y_i = x \right) \mathbb{P}_\theta(\tilde{N}(C_j) = q) \\ &= \tilde{\eta}(c_j) \tilde{\nu} Leb(C_j) e^{-\tilde{\nu} Leb(C_j)} + o\left(Leb(C_j) \right), \end{aligned}$$

as $n \to \infty$. This proves the second part of the lemma.

(Proof of first part) As for the first part, use Equations (S1) and (S2) to write

$$\begin{aligned}\prod_{j=1}^{4}\mathbb{P}_\theta\left(L(C_j)=0\right) &= e^{-\tilde{\nu}\sum_{j=1}^{4}Leb(C_j)}+o\left(\sum_{j=1}^{4}Leb(C_j)\right)\\ &= e^{-\tilde{\nu}Leb(A_{n\Delta}\cap A_{k\Delta})}+o\left(Leb(A_{n\Delta}\cap A_{k\Delta})\right)\\ &= 1-\tilde{\nu}Leb(A_{n\Delta}\cap A_{k\Delta})+o\left(Leb(A_{n\Delta}\cap A_{k\Delta})\right),\end{aligned}$$

as $n\to\infty$, where we in the last line Taylor expanded the exponential function. This proves the first part of the lemma.

□

Proof of Lemma 4.1. Using Bayes' theorem and the unconditional independence of $L(A_t\cap A_{t+h})$ and $L(A_t\setminus A_{t+h})$, we have for $x\in\mathbb{N}\cup\{0\}$ and $l\in\{0,1,\ldots,x\}$:

$$\begin{aligned}&\mathbb{P}(L(A_t\cap A_{t+h})=l|X_t=x)=\frac{\mathbb{P}(X_t=x|L(A_t\cap A_{t+h})=l)\mathbb{P}(L(A_t\cap A_{t+h})=l)}{\mathbb{P}(X_t=x)}\\ &=\frac{\mathbb{P}(L(A_t\cap A_{t+h})+L(A_t\setminus A_{t+h})=x|L(A_t\cap A_{t+h})=l)\mathbb{P}(L(A_t\cap A_{t+h})=l)}{\mathbb{P}(X_t=x)}\\ &=\frac{\mathbb{P}(L(A_t\setminus A_{t+h})=x-l|L(A_t\cap A_{t+h})=l)\mathbb{P}(L(A_t\cap A_{t+h})=l)}{\mathbb{P}(X_t=x)}\\ &=\frac{\mathbb{P}(L(A_t\setminus A_{t+h})=x-l)\mathbb{P}(L(A_t\cap A_{t+h})=l)}{\mathbb{P}(X_t=x)}.\end{aligned}$$

□

Proof of Proposition 4.1. Using the conditional law of total probability we obtain the following convolution formula

$$\begin{aligned}&\mathbb{P}(X_{t+h}=x_{t+h}|X_t=x_t)=\mathbb{P}(L(A_t\cap A_{t+h})+L(A_{t+h}\setminus A_t)=x_{t+h}|X_t=x_t)\\ &=\sum_{c=0}^{\min(x_t,x_{t+h})}\mathbb{P}(L(A_t\cap A_{t+h})+L(A_{t+h}\setminus A_t)=x_{t+h}|X_t=x_t,L(A_t\cap A_{t+h})=c)\\ &\cdot\mathbb{P}(L(A_t\cap A_{t+h})=c|X_t=x_t)\\ &=\sum_{c=0}^{\min(x_t,x_{t+h})}\mathbb{P}(L(A_{t+h}\setminus A_t)=x_{t+h}-c|X_t=x_t,L(A_t\cap A_{t+h})=c)\mathbb{P}(L(A_t\cap A_{t+h})=c|X_t=x_t)\\ &=\sum_{c=0}^{\min(x_t,x_{t+h})}\mathbb{P}(L(A_{t+h}\setminus A_t)=x_{t+h}-c)\mathbb{P}(L(A_t\cap A_{t+h})=c|X_t=x_t).\end{aligned}$$

□

Proof of Proposition C.1. Ignoring the dependence on θ we have

$$f(x_t,x_s)=\sum_{c=-\infty}^{\infty}f(x_t,x_s|c)f(c)=\sum_{c=-\infty}^{\infty}f(x_t|x_s,c)f(x_s|c)f(c).$$

Since, conditionally on $L(A_t \cap A_s) = c$, X_t and X_s are independent we have $f(x_t|x_s, c) = f(x_t|c)$ and thus

$$f(x_t, x_s) = \sum_{c=-\infty}^{\infty} f(x_t|c) f(x_s|c) f(c),$$

which shows that sampling from $f(c; \theta)$ delivers the quantity we need. The rest of the proposition is obvious. □

3 Additional simulation results

3.1 Simulation results supplementing those from the main paper

The parameter values used in the simulation studies are given in Table S1 and the implied marginal distributions and autocorrelation structures are shown in Figure S2. This table and this figure were also shown in the main paper, but we reproduce them here for completeness.

The simulation results, for various values of n, are shown in Tables S2–S7 for the six DGPs of Table S1. We report the median, the median bias, and the root median squared error (RMSE) of the estimator, calculated over the 500 Monte Carlo replications. The reason for reporting the median, instead of the mean, is that we found that when the number of observations, n, is small, the estimation approach will occasionally result in large outliers in few of the Monte Carlo runs, thus skewing the results (this was the case for both the MCL and GMM estimators).

From the tables, we see evidence of the MCL estimator being consistent, i.e. the bias converges towards zero as the number of observations, n, grows. As expected, the estimator is most precise for the simpler models, e.g. the Poisson-Exp IVT (Table S2) and somewhat less precise for the more complex models, e.g. the NB-IG model (Table S6) and NB-Gamma model (Table S7).

3.1.1 Finite sample properties of the model selection procedure

This section illustrates the use, and finite sample properties, of the model selection procedure introduced in Section 3.1.3 of the main paper. Consider $n = 4000$ equidistant observations of an IVT process on a grid of $\Delta = 0.10$, i.e. $X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$.[14] For each of the six possible models, we then calculate the three goodness-of-fit measures, namely the value of the maximized composite likelihood function CL, the AIC-like composite likelihood information criteria $CLAIC$, and the BIC-like composite likelihood information criteria $CLBIC$. The model which has the maximum value of a criteria is "selected" by that criteria. We repeat this process for 100 Monte Carlo replications and the six different DGPs using the parameters of Table S1. Figure S3 reports the "selection rates" of the models, i.e. the fraction of times that a model, given on the x-axis, is selected, for each of the three different criteria. Each panel in the figure corresponds to a particular DGP, as shown above the respective panels.

Consider, for instance, the case where the true DGP is the NB-Exp IVT model. The results from using this DGP are given in the upper right panel of Figure S3. In this case, when we estimate the six different models and calculate the three goodness-of-fit measures, the true model (i.e. NB-Exp) has the highest composite likelihood value in 70% of simulations. In contrast, the CLAIC and CLBIC result in selecting the true model in 73% and 81% of the simulations, respectively. Note that since the models considered here are not nested, it is not necessarily the case that the

[14] The number of simulated observations, $n = 4000$, the space between observations, $\Delta = 0.10$, and the tuning parameter, $K = 10$, are chosen such as to be comparable to the data studied in the empirical section of the main paper.

maximized composite likelihood value CL will be larger for the more complicated models.

Overall, Figure S3 indicates that the model selection procedure is quite accurate when the marginal distribution of the DGP is the Negative Binomial distribution. Conversely, when the marginal distribution of the DGP is the Poisson distribution, the correct model is chosen less often. However, in these situations, it is often the case that although the Negative Binomial distribution is (incorrectly) preferred to the Poisson distribution, the correct trawl function (autocorrelation structure) is nonetheless selected.

Lastly, to examine the effect of the tuning parameter K on the model selection procedure, we ran the same experiment but using both $K = 5$ and $K = 20$ (results not shown here, but available upon request). We find that the model selection procedure deteriorates when $K = 5$, while it performs similarly to that shown in Figure S3 when $K = 20$, indicating that it is important to set K sufficiently large value, so that the selection criteria can properly distinguish between the models.

Table S1: *Parameter values used in simulation studies*

DGP	ν	m	p	λ	δ	γ	H	α
P-Exp	17.50			1.80				
P-IG	17.50				1.80	0.80		
P-Gamma	17.50						1.70	0.80
NB-Exp		7.50	0.70	1.80				
NB-IG		7.50	0.70		1.80	0.80		
NB-Gamma		7.50	0.70				1.70	0.80

Parameter values for the six different DGPs used in the simulation studies of Section 3.1. The value $\nu = mp/(1-p)$ with $m = 7.5$ and $p = 0.70$ is chosen such that the first moment of the Poisson and Negative Binomial Lévy bases are matched. Marginal distributions and autocorrelation functions implied by these parameter values are shown in Figure S2.

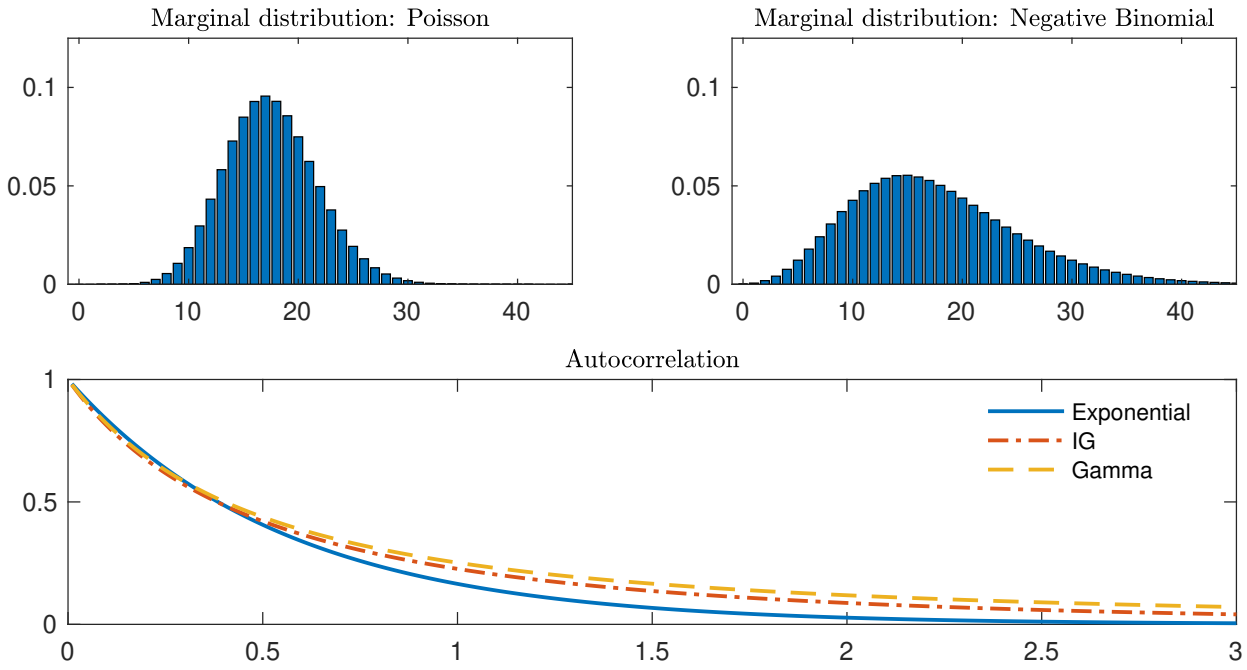

Figure S2: *Marginal distributions of the Lévy bases (top) and autocorrelations of the DGPs (bottom) used in the simulation studies of Section 3.1. The marginal distribution and autocorrelation structure of IVT processes can be specified independently, resulting in six different DGPs in this setup (P-Exp, P-IG, P-Gamma, NB-Exp, NB-IG, NB-Gamma). The parameter values used to produce the plots are given in Table S1.*

Table S2: *MCL estimation results: Poisson trawl process with exponential trawl function*

	$\hat{\nu}$ ($\nu = 17.5$)			$\hat{\lambda}$ ($\lambda = 1.8$)		
n	Med.	Bias	RMSE	Med.	Bias	RMSE
100	17.3213	−0.1787	1.8616	1.8351	0.0351	0.2035
250	17.4204	−0.0796	1.1761	1.8009	0.0009	0.1279
500	17.5179	0.0179	0.8444	1.8012	0.0012	0.0883
1000	17.6023	0.1023	0.6164	1.8049	0.0049	0.0626
2000	17.5540	0.0540	0.4538	1.8026	0.0026	0.0476
4000	17.5099	0.0099	0.3038	1.8000	0.0000	0.0327
8000	17.5302	0.0302	0.2197	1.8036	0.0036	0.0222

Median (Med.), median bias (Bias) and root median squared error (RMSE) of the MCL estimator with $K = 1$. *DGP: Poisson-Exponential IVT process. The IVT process* X_t *is simulated on the grid* $t = \Delta, 2\Delta, \ldots, n\Delta$, *with* $\Delta = 0.10$, *see Table S1 for the values of the parameters used in the simulations. Number of Monte Carlo simulations:* 500.

Table S3: *CL estimation results: Poisson trawl process with IG trawl function*

	$\hat{\nu}$ ($\nu = 17.5$)			$\hat{\delta}$ ($\delta = 1.8$)			$\hat{\gamma}$ ($\gamma = 0.8$)		
n	Med.	Bias	RMSE	Med.	Bias	RMSE	Med.	Bias	RMSE
250	17.7984	0.2984	1.8764	2.4177	0.6177	0.9923	1.0985	0.2985	0.5365
500	17.7749	0.2749	1.4921	2.0205	0.2205	0.6317	0.9306	0.1306	0.3352
1000	17.6514	0.1514	1.0344	1.9141	0.1141	0.4896	0.8592	0.0592	0.2422
2000	17.5182	0.0182	0.8090	1.8977	0.0977	0.3413	0.8333	0.0333	0.1864
4000	17.5273	0.0273	0.5931	1.8120	0.0120	0.2426	0.8051	0.0051	0.1347
8000	17.4966	−0.0034	0.4125	1.8072	0.0072	0.1692	0.8030	0.0030	0.0842

Median (Med.), median bias (Bias) and root median squared error (RMSE) of the MCL estimator with $K = 10$. *DGP: Poisson-IG IVT. The IVT process* X_t *is simulated on the grid* $t = \Delta, 2\Delta, \ldots, n\Delta$, *with* $\Delta = 0.10$, *see Table S1 for the values of the parameters used in the simulations. Number of Monte Carlo simulations:* 500.

Table S4: *CL estimation results: Poisson trawl process with* Γ *trawl function*

	$\hat{\nu}$ ($\nu = 17.5$)			$\hat{H}$ ($H = 1.7$)			$\hat{\alpha}$ ($\alpha = 0.8$)		
n	Med.	Bias	RMSE	Med.	Bias	RMSE	Med.	Bias	RMSE
250	17.8207	0.3207	1.8470	3.0719	1.3719	1.4020	1.5159	0.7159	0.7302
500	17.5946	0.0946	1.2180	2.1831	0.4831	0.9537	1.0463	0.2463	0.4816
1000	17.4396	−0.0604	0.9877	1.8563	0.1563	0.7227	0.8914	0.0914	0.3636
2000	17.4540	−0.0460	0.6642	1.7807	0.0807	0.5371	0.8406	0.0406	0.2728
4000	17.5048	0.0048	0.4821	1.6902	−0.0098	0.3851	0.7784	−0.0216	0.2004
8000	17.5612	0.0612	0.3925	1.6433	−0.0567	0.2464	0.7607	−0.0393	0.1189

Median (Med.), median bias (Bias) and root median squared error (RMSE) of the MCL estimator with $K = 10$. *DGP: Poisson-Gamma IVT. The IVT process* X_t *is simulated on the grid* $t = \Delta, 2\Delta, \ldots, n\Delta$, *with* $\Delta = 0.10$, *see Table S1 for the values of the parameters used in the simulations. Number of Monte Carlo simulations:* 500.

Table S5: *CL estimation results: NB trawl process with exponential trawl function*

	$\hat{m}$ ($m = 7.5$)			$\hat{p}$ ($p = 0.7$)			$\hat{\lambda}$ ($\lambda = 1.8$)		
n	Med.	Bias	RMSE	Med.	Bias	RMSE	Med.	Bias	RMSE
100	8.9761	1.4761	2.3850	0.6649	−0.0351	0.0623	1.9197	0.1197	0.2595
250	7.9680	0.4680	1.3264	0.6872	−0.0128	0.0353	1.8311	0.0311	0.1431
500	7.8443	0.3443	0.9956	0.6896	−0.0104	0.0287	1.8202	0.0202	0.1010
1000	7.6891	0.1891	0.6964	0.6970	−0.0030	0.0202	1.8106	0.0106	0.0728
2000	7.5855	0.0855	0.5010	0.6972	−0.0028	0.0143	1.8001	0.0001	0.0470
4000	7.5362	0.0362	0.3484	0.6993	−0.0007	0.0095	1.8013	0.0013	0.0316
8000	7.5265	0.0265	0.2365	0.6993	−0.0007	0.0065	1.7994	−0.0006	0.0251

Median (Med.), median bias (Bias) and root median squared error (RMSE) of the MCL estimator with $K = 1$. *DGP: Negative Binomial-Exponential IVT process. The IVT process* X_t *is simulated on the grid* $t = \Delta, 2\Delta, \ldots, n\Delta$, *with* $\Delta = 0.10$, *see Table S1 for the values of the parameters used in the simulations. Number of Monte Carlo simulations:* 500.

Table S6: *CL estimation results: NB trawl process with IG trawl function*

	$\hat{m}$ ($m = 7.5$)			$\hat{p}$ ($p = 0.7$)			$\hat{\delta}$ ($\delta = 1.8$)			$\hat{\gamma}$ ($\gamma = 0.8$)		
n	Med.	Bias	RMSE	Med.	Bias	RMSE	Med.	Bias	RMSE	Med.	Bias	RMSE
250	9.4842	1.9842	2.2867	0.6695	−0.0305	0.0501	3.1478	1.3478	1.3540	1.3513	0.5513	0.6440
500	8.1866	0.6866	1.4351	0.6885	−0.0115	0.0335	2.3665	0.5665	0.8803	1.0506	0.2506	0.4657
1000	7.7323	0.2323	0.8607	0.6941	−0.0059	0.0256	1.9682	0.1682	0.6219	0.8834	0.0834	0.3099
2000	7.6150	0.1150	0.5755	0.6970	−0.0030	0.0176	1.9398	0.1398	0.4398	0.8588	0.0588	0.2116
4000	7.5561	0.0561	0.4348	0.6994	−0.0006	0.0129	1.8866	0.0866	0.3547	0.8351	0.0351	0.1741
8000	7.5317	0.0317	0.3103	0.6990	−0.0010	0.0088	1.8470	0.0470	0.2266	0.8247	0.0247	0.1063

Median (Med.), median bias (Bias) and root median squared error (RMSE) of the MCL estimator with $K = 10$. *DGP: Negative Binomial-IG IVT. The IVT process* X_t *is simulated on the grid* $t = \Delta, 2\Delta, \ldots, n\Delta$, *with* $\Delta = 0.10$, *see Table S1 for the values of the parameters used in the simulations. Number of Monte Carlo simulations:* 500.

Table S7: *CL estimation results: NB trawl process with* Γ *trawl function*

	$\hat{m}$ ($m = 7.5$)			$\hat{p}$ ($p = 0.7$)			$\hat{H}$ ($H = 1.7$)			$\hat{\alpha}$ ($\alpha = 0.8$)		
n	Med.	Bias	RMSE	Med.	Bias	RMSE	Med.	Bias	RMSE	Med.	Bias	RMSE
250	9.2542	1.7542	2.3233	0.6678	−0.0322	0.0532	4.0968	2.3968	2.3968	1.9812	1.1812	1.1814
500	8.3098	0.8098	1.5008	0.6799	−0.0201	0.0384	2.6049	0.9049	1.2064	1.2160	0.4160	0.5768
1000	7.9214	0.4214	0.9728	0.6890	−0.0110	0.0257	2.2420	0.5420	0.9221	1.0470	0.2470	0.4404
2000	7.7270	0.2270	0.6394	0.6938	−0.0062	0.0179	1.9538	0.2538	0.7091	0.9244	0.1244	0.3413
4000	7.5876	0.0876	0.4594	0.6977	−0.0023	0.0130	1.8234	0.1234	0.5016	0.8545	0.0545	0.2418
8000	7.5249	0.0249	0.3355	0.6990	−0.0010	0.0091	1.7745	0.0745	0.3466	0.8436	0.0436	0.1671

Median (Med.), median bias (Bias) and root median squared error (RMSE) of the MCL estimator with $K = 10$. *DGP: Negative Binomial-Gamma IVT process. The IVT process* X_t *is simulated on the grid* $t = \Delta, 2\Delta, \ldots, n\Delta$, *with* $\Delta = 0.10$, *see Table S1 for the values of the parameters used in the simulations. Number of Monte Carlo simulations:* 500.

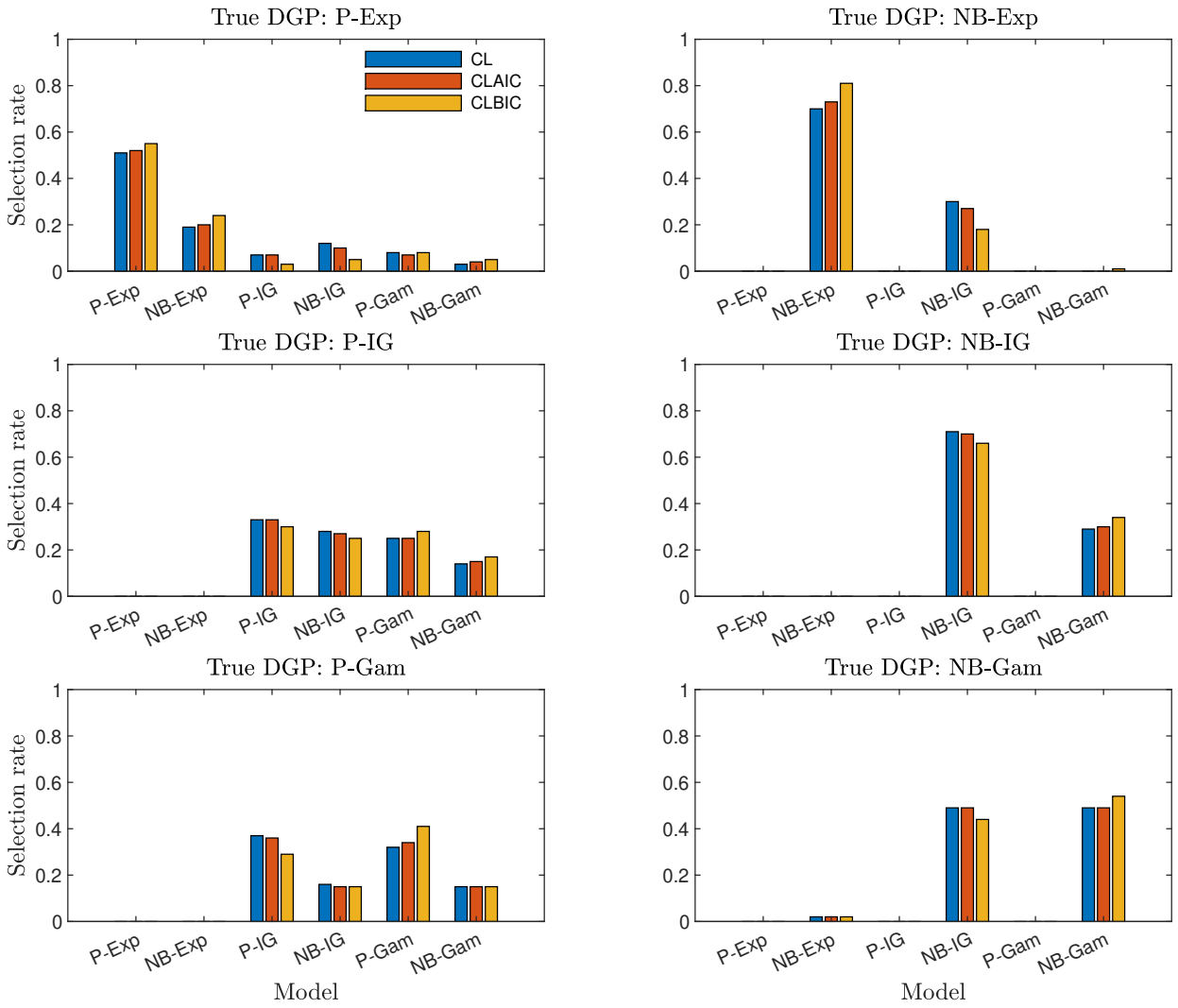

Figure S3: *Simulation study of model selection procedure. Each plot represents the outcome of a separate Monte Carlo study, where the true DGP in the study is given above the plot. The numbers plotted are average selection rates of the models given on the x-axis, using a given criteria over* $M = 100$ *Monte Carlo simulations. For each Monte Carlo replication,* $n = 4000$ *observations of the true DGP are simulated on a grid with step size* $\Delta = 0.10$*. The parameters used in the study are given in Table S1 and we set* $K = 10$*.*

3.2 Alternative simulation setup

We perform simulation experiments similar to those in the main paper but with a different set of simulation settings. The parameter values used for the DGPs in this study are given in Table S8; the associated implied marginal distributions of the underlying Lévy bases and autocorrelations of the IVT processes are shown in Figure S4.

In this simulation study, we simulate n observations of an IVT process X_t on an equidistant grid of size $\Delta = 0.10$. For the IVTs based on the exponential trawl function, we set $K = 1$, while we set $K = 3$ for the remaining IVTs. This should be contrasted to the setup of the main paper, where we set $K = 10$. The finite sample estimation results can be found in Tables S9–S14. Figure S5 plots the relative RMSE of the MCL estimator compared to the MM estimator; numbers smaller than one favor the MCL estimator.

Table S8: *Parameter values used in simulation setup 2*

DGP	ν	m	p	λ	δ	γ	H	α
P-Exp	5.00			1.00				
P-IG	5.00				0.75	0.50		
P-Gamma	5.00						0.50	0.75
NB-Exp		2.14	0.70	1.00				
NB-IG		2.14	0.70		0.75	0.50		
NB-Gamma		2.14	0.70				0.50	0.75

Parameter values for the six different DGPs used in the simulation studies of the Supplementary Material. See the various Examples of the main paper for details. The value $m = \nu(1-p)/p$ with $\nu = 5$ is chosen such that the first moment of the Poisson and Negative Binomial Lévy bases are matched. Marginal distributions and autocorrelation function implied by these parameter values are shown in Figure S4.

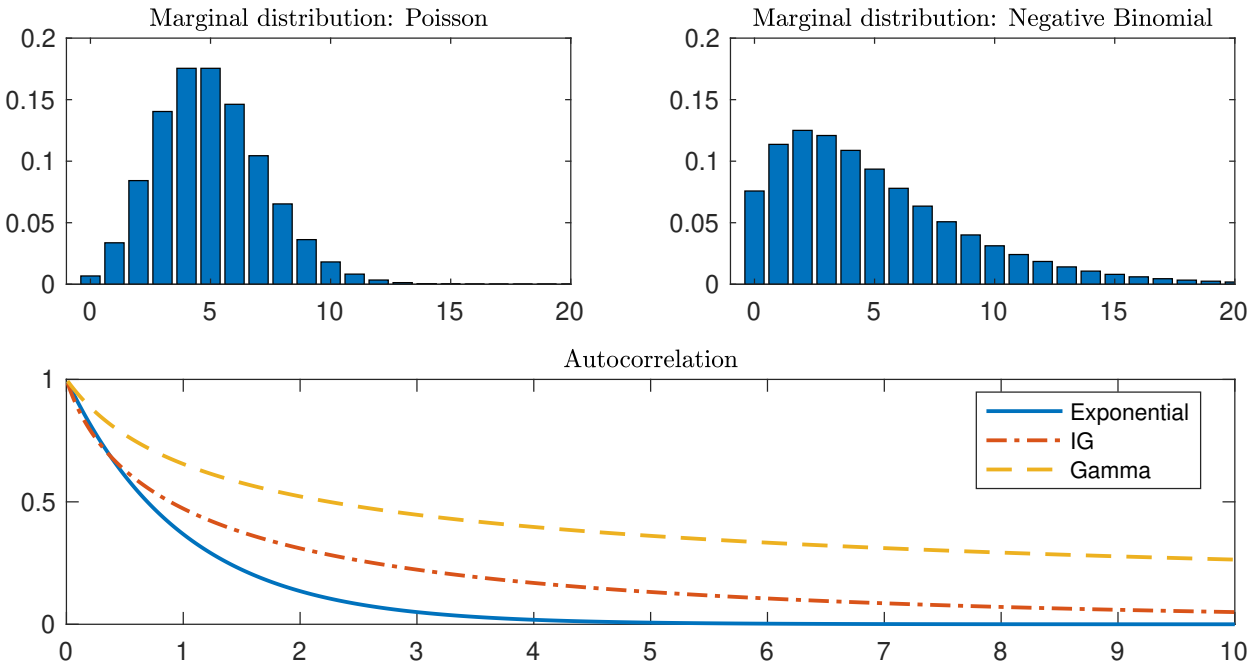

Figure S4: *Marginal distributions of the Lévy bases and autocorrelations of the DGPs used in the simulation studies of Section 3. The marginal distribution and autocorrelation structure of IVT processes can be specified independently, resulting in six different DGPs in this setup (P-Exp, P-IG, P-Gamma, NB-Exp, NB-IG, NB-Gamma). The parameter values used to produce the plots are given in Table S8. Note that the marginal distribution shown in the top plots are of the underlying Lévy bases and not of the IVT process itself.*

Table S9: *CL estimation results: Poisson trawl process with exponential trawl function*

	$\hat{\nu}$ ($\nu = 5$)			$\hat{\lambda}$ ($\lambda = 1$)		
Nobs	Avg	Bias	RMSE	Avg	Bias	RMSE
100	4.9994	−0.0006	0.6464	1.0131	0.0131	0.1220
250	5.0345	0.0345	0.4055	1.0099	0.0099	0.0756
500	5.0468	0.0468	0.2908	1.0074	0.0074	0.0623
1000	5.0100	0.0100	0.2181	1.0067	0.0067	0.0441
2000	5.0217	0.0217	0.1383	1.0026	0.0026	0.0288
4000	5.0103	0.0103	0.1058	1.0009	0.0009	0.0203
8000	5.0012	0.0012	0.0722	1.0005	0.0005	0.0150

Median (Med.), median bias (Bias) and root median squared error (RMSE) of the MCL estimator. DGP: Poisson-Exponential IVT process. The IVT process X_t is simulated on the grid $t = \Delta, 2\Delta, \ldots, n\Delta$, with $\Delta = 0.10$, see Table S8 for the values of the parameters used in the simulations. $K = 1$. Number of Monte Carlo simulations: 500.

Table S10: *CL estimation results: Poisson trawl process with IG trawl function*

	$\hat{\nu}$ ($\nu = 5$)			$\hat{\delta}$ ($\delta = 0.75$)			$\hat{\gamma}$ ($\gamma = 0.5$)		
Nobs	Avg	Bias	RMSE	Avg	Bias	RMSE	Avg	Bias	RMSE
250	4.8852	−0.1148	0.6536	0.8535	0.1035	0.2654	5.6836	5.1836	31.0800
500	4.8907	−0.1093	0.4725	0.8241	0.0741	0.1833	1.1475	0.6475	6.2198
1000	4.9770	−0.0230	0.3128	0.7849	0.0349	0.1266	0.5699	0.0699	0.2729
2000	4.9789	−0.0211	0.2322	0.7641	0.0141	0.0890	0.5281	0.0281	0.1196
4000	4.9921	−0.0079	0.1516	0.7525	0.0025	0.0640	0.5118	0.0118	0.0773
8000	4.9997	−0.0003	0.1203	0.7533	0.0033	0.0421	0.5042	0.0042	0.0543

Median (Med.), median bias (Bias) and root median squared error (RMSE) of the MCL estimator. DGP: Poisson-IG IVT. The IVT process X_t is simulated on the grid $t = \Delta, 2\Delta, \ldots, n\Delta$, with $\Delta = 0.10$, see Table S8 for the values of the parameters used in the simulations. $K = 3$. Number of Monte Carlo simulations: 500.

Table S11: *CL estimation results: Poisson trawl process with Γ trawl function*

	$\hat{\nu}$ ($\nu = 5$)			$\hat{H}$ ($H = 0.5$)			$\hat{\alpha}$ ($\alpha = 0.75$)		
Nobs	Avg	Bias	RMSE	Avg	Bias	RMSE	Avg	Bias	RMSE
250	5.0958	0.0958	0.4897	0.6691	0.1691	0.3452	0.9620	0.2120	0.5434
500	4.9864	−0.0136	0.3323	0.6360	0.1360	0.2879	0.9710	0.2210	0.4283
1000	4.9781	−0.0219	0.2526	0.5957	0.0957	0.2109	0.8667	0.1167	0.3183
2000	4.9866	−0.0134	0.1776	0.5765	0.0765	0.1531	0.8480	0.0980	0.2443
4000	4.9789	−0.0211	0.1233	0.5519	0.0519	0.1156	0.8223	0.0723	0.1678
8000	4.9720	−0.0280	0.0945	0.5515	0.0515	0.0864	0.8147	0.0647	0.1234

Median (Med.), median bias (Bias) and root median squared error (RMSE) of the MCL estimator. DGP: Poisson-Gamma IVT. The IVT process X_t is simulated on the grid $t = \Delta, 2\Delta, \ldots, n\Delta$, with $\Delta = 0.10$, see Table S8 for the values of the parameters used in the simulations. $K = 3$. Number of Monte Carlo simulations: 500.

Table S12: *CL estimation results: NB trawl process with exponential trawl function*

	$\hat{m}$ ($m = 2.1429$)			$\hat{p}$ ($p = 0.7$)			$\hat{\lambda}$ ($\lambda = 1$)		
Nobs	Avg	Bias	RMSE	Avg	Bias	RMSE	Avg	Bias	RMSE
100	2.7614	0.6186	0.8445	0.6268	−0.0732	0.1010	1.0021	0.0021	0.1902
250	2.3957	0.2528	0.4848	0.6692	−0.0308	0.0618	0.9921	−0.0079	0.1206
500	2.1957	0.0529	0.3009	0.6883	−0.0117	0.0394	1.0038	0.0038	0.0791
1000	2.2200	0.0771	0.2357	0.6912	−0.0088	0.0309	1.0008	0.0008	0.0596
2000	2.1887	0.0458	0.1632	0.6935	−0.0065	0.0215	1.0030	0.0030	0.0410
4000	2.1630	0.0201	0.1126	0.6967	−0.0033	0.0144	0.9999	−0.0001	0.0273
8000	2.1557	0.0128	0.0805	0.6985	−0.0015	0.0109	1.0012	0.0012	0.0210

Median (Med.), median bias (Bias) and root median squared error (RMSE) of the MCL estimator. DGP: Negative Binomial-Exponential IVT process. The IVT process X_t is simulated on the grid $t = \Delta, 2\Delta, \ldots, n\Delta$, with $\Delta = 0.10$, see Table S8 for the values of the parameters used in the simulations. $K = 1$. Number of Monte Carlo simulations: 500.

Table S13: *CL estimation results: NB trawl process with IG trawl function*

	$\hat{m}$ ($m = 2.1429$)			$\hat{p}$ ($p = 0.7$)			$\hat{\delta}$ ($\delta = 0.75$)			$\hat{\gamma}$ ($\gamma = 0.5$)		
Nobs	Avg	Bias	RMSE	Avg	Bias	RMSE	Avg	Bias	RMSE	Avg	Bias	RMSE
250	2.5049	0.3621	0.5605	0.6361	−0.0639	0.0827	0.9671	0.2171	0.3323	0.6664	0.1664	0.2521
500	2.3485	0.2056	0.4236	0.6688	−0.0312	0.0632	0.8739	0.1239	0.2095	0.5738	0.0738	0.1402
1000	2.2625	0.1197	0.3055	0.6767	−0.0233	0.0450	0.7991	0.0491	0.1460	0.5280	0.0280	0.0986
2000	2.1975	0.0547	0.2073	0.6937	−0.0063	0.0303	0.7757	0.0257	0.1099	0.5147	0.0147	0.0769
4000	2.1757	0.0328	0.1384	0.6936	−0.0064	0.0199	0.7777	0.0277	0.0800	0.5203	0.0203	0.0573
8000	2.1661	0.0232	0.0970	0.6978	−0.0022	0.0152	0.7631	0.0131	0.0560	0.5127	0.0127	0.0380

Median (Med.), median bias (Bias) and root median squared error (RMSE) of the MCL estimator. DGP: Negative Binomial-IG IVT. The IVT process X_t is simulated on the grid $t = \Delta, 2\Delta, \ldots, n\Delta$, with $\Delta = 0.10$, see Table S8 for the values of the parameters used in the simulations. $K = 3$. Number of Monte Carlo simulations: 500.

Table S14: *CL estimation results: NB trawl process with Γ trawl function*

	$\hat{m}$ ($m = 2.1429$)			$\hat{p}$ ($p = 0.7$)			$\hat{H}$ ($H = 0.5$)			$\hat{\alpha}$ ($\alpha = 0.75$)		
Nobs	Avg	Bias	RMSE	Avg	Bias	RMSE	Avg	Bias	RMSE	Avg	Bias	RMSE
500	2.6635	0.5206	0.5684	0.6240	−0.0760	0.0873	3.0978	2.5978	2.5978	4.6910	3.9410	3.9414
1000	2.4640	0.3212	0.4184	0.6510	−0.0490	0.0608	1.1213	0.6213	0.6213	1.6475	0.8975	0.8975
2000	2.3565	0.2136	0.3091	0.6659	−0.0341	0.0461	0.7854	0.2854	0.2887	1.1460	0.3960	0.4207
4000	2.2840	0.1411	0.1944	0.6739	−0.0261	0.0327	0.7204	0.2204	0.2391	1.0852	0.3352	0.3518
8000	2.2302	0.0873	0.1583	0.6843	−0.0157	0.0256	0.7034	0.2034	0.2142	1.0659	0.3159	0.3259

Median (Med.), median bias (Bias) and root median squared error (RMSE) of the MCL estimator. DGP: Negative Binomial-Gamma IVT process. The IVT process X_t is simulated on the grid $t = \Delta, 2\Delta, \ldots, n\Delta$, with $\Delta = 0.10$, see Table S8 for the values of the parameters used in the simulations. $K = 3$. Number of Monte Carlo simulations: 500.

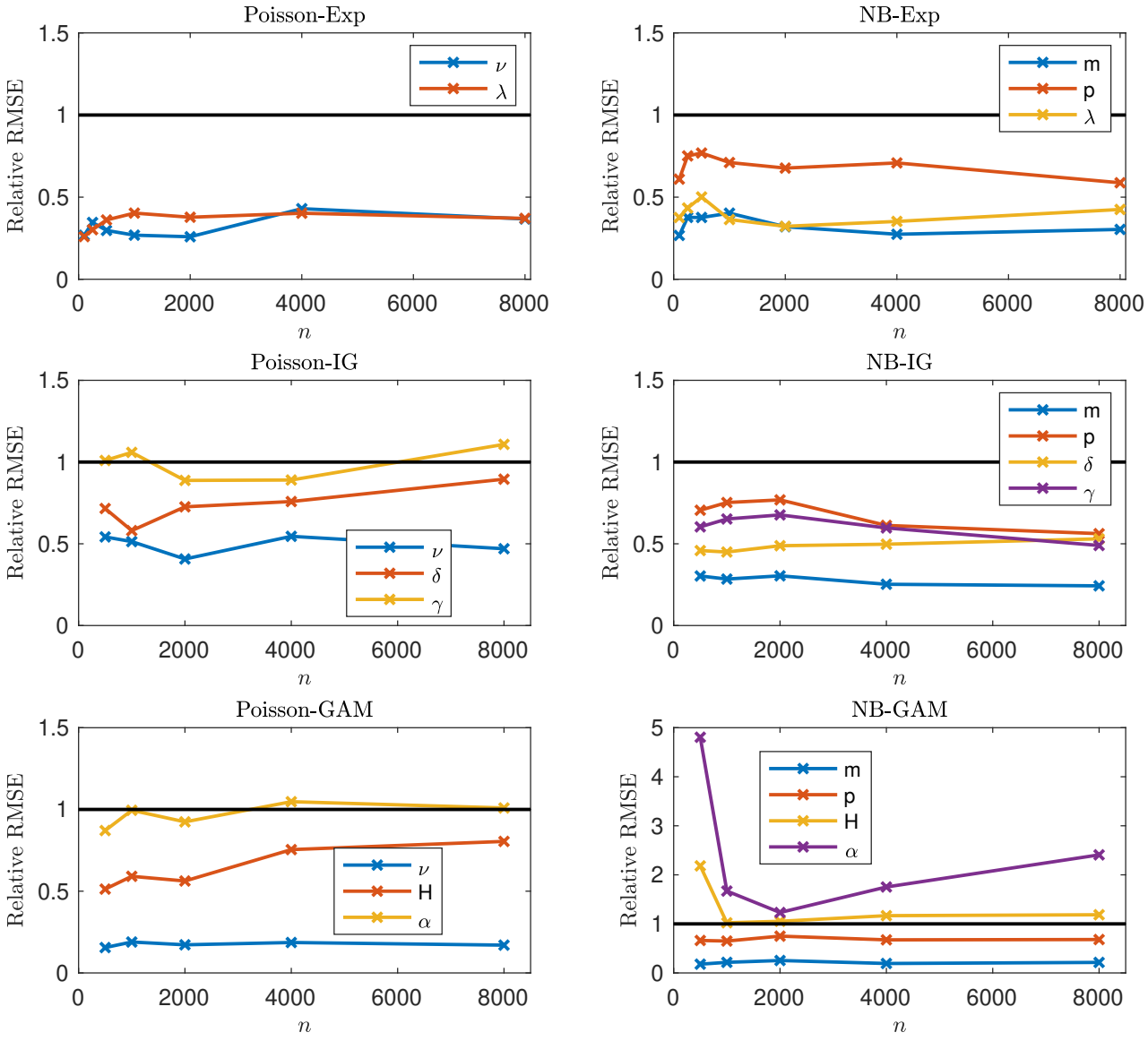

Figure S5: *Root median square error (RMSE) of the MCL estimator divided by the RMSE of the MM estimator. The underlying IVT process X_t is simulated on the grid $t = \Delta, 2\Delta, \ldots, n\Delta$, with $\Delta = 0.10$, see Table S8 for the values of the parameters used in the simulations. For the Poisson-Exp and NB-Exp we set $K = 1$; for the other DGPs we set $K = 3$. We also conducted the comparison with $K = 5$, as suggested in Barndorff-Nielsen et al. (2014) with similar results (results not presented here but available from the authors upon request).*

4 Additional forecasting results

Figures S6–S8 report forecasting results analogous to those of Section 6.1 in the main paper, but now using the conditional mean, instead of the conditional mode, as a point forecast. That is, using the notation of Section 6.1 in the main paper, to construct a conditional point forecast $\hat{x}_{i|i-1}$, we set $\hat{x}_{i|i-h} = \sum_{k=0}^{M} \widehat{\mathbb{P}}(X_{i|i-h} = k)k$, where $M \geq 1$ is a large (cut-off) number and $\widehat{\mathbb{P}}$ is the estimated predictive density of the IVT model. In our data, the maximum spread level in the in-sample period was 39 and we consequently we set $M = 60$ such that we are comfortable that we will have $x_i \leq M$ for all i.

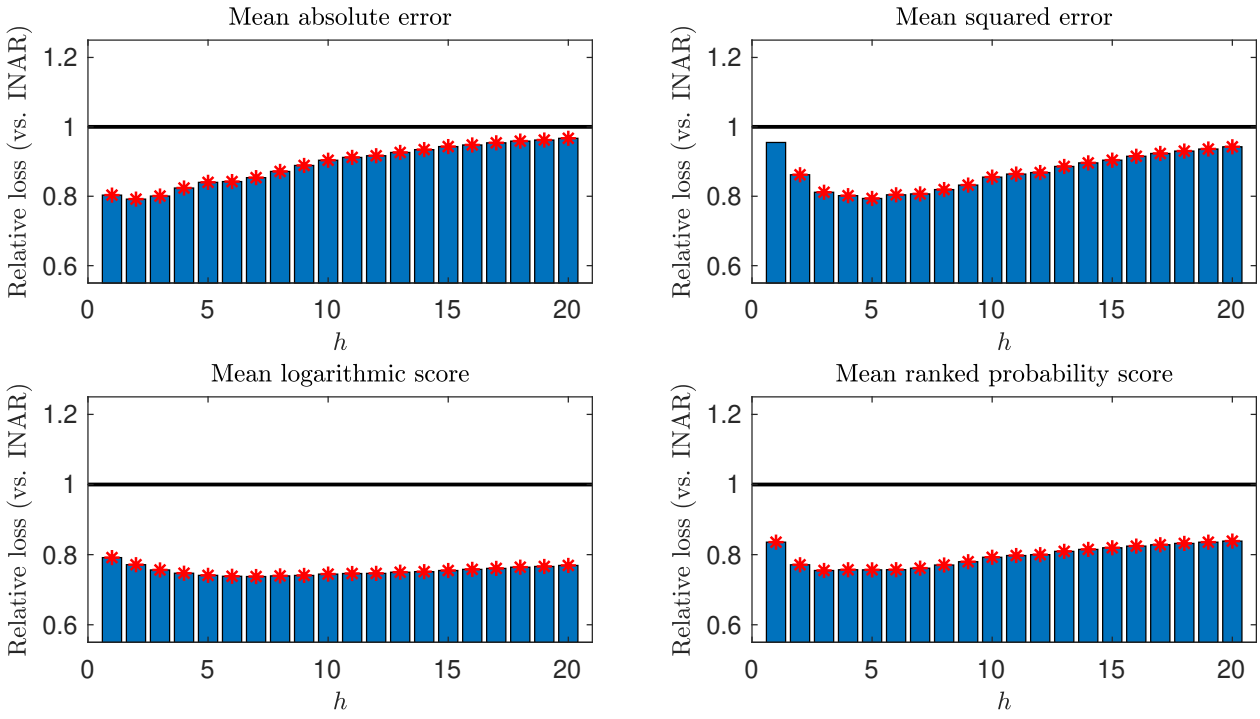

Figure S6: *Forecasting the spread level of the A stock on May 4, 2020. Four different loss metrics and twenty forecast horizons,* $h = 1, 2, \ldots, 20$. *The numbers plotted are relative average losses of the NB-Gamma forecasting model, compared with the Poissonian INAR(1) model, over* $n_{oos} = 720$ *out-of-sample forecasts. A circle above the bars indicates rejection null of equal forecasting performance between the two models, against the alternative that the NB-IG model provides superior forecasts, using the Diebold-Mariano (Diebold and Mariano, 1995) test at a* 5% *level; an asterisk denotes rejection at a* 1% *level.*

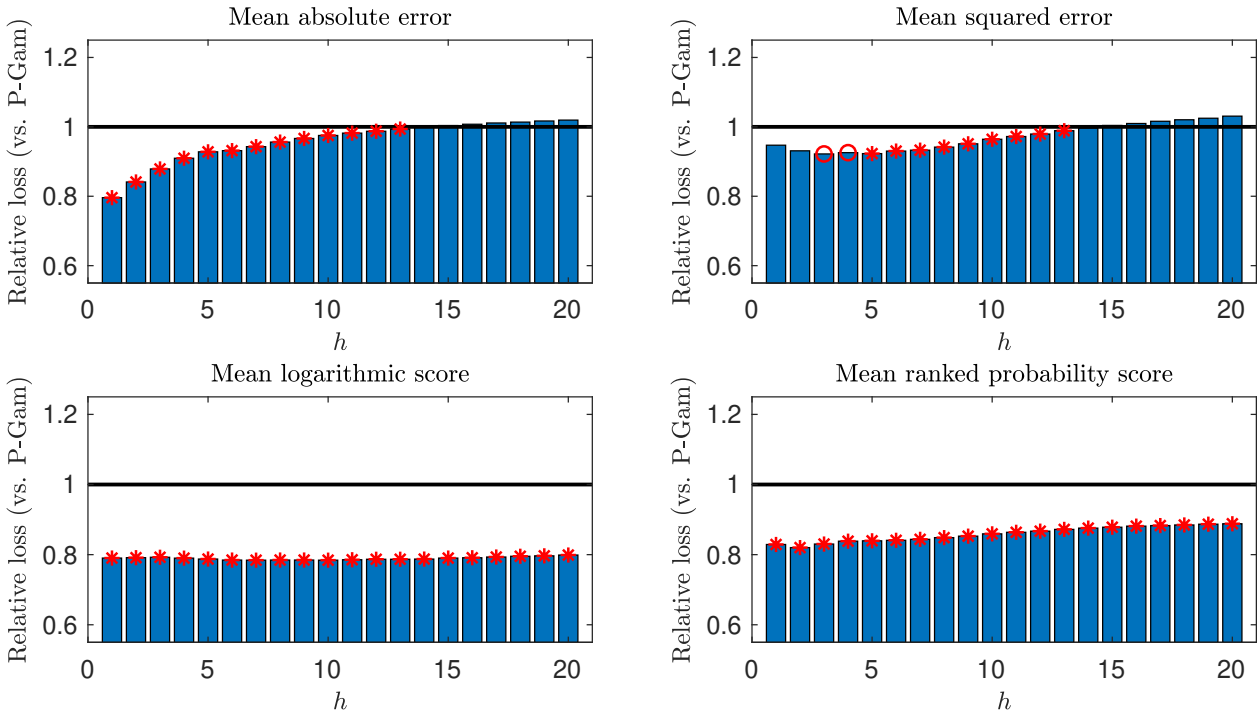

Figure S7: *Forecasting the spread level of the A stock on May 4, 2020. Four different loss metrics and twenty forecast horizons,* $h = 1, 2, \ldots, 20$. *The numbers plotted are relative average losses of the NB-Gamma forecasting model, compared with the Poisson-Gamma model, over* $n_{oos} = 720$ *out-of-sample forecasts. A circle above the bars indicates rejection null of equal forecasting performance between the two models, against the alternative that the NB-IG model provides superior forecasts, using the Diebold-Mariano (Diebold and Mariano, 1995) test at a* 5% *level; an asterisk denotes rejection at a* 1% *level.*

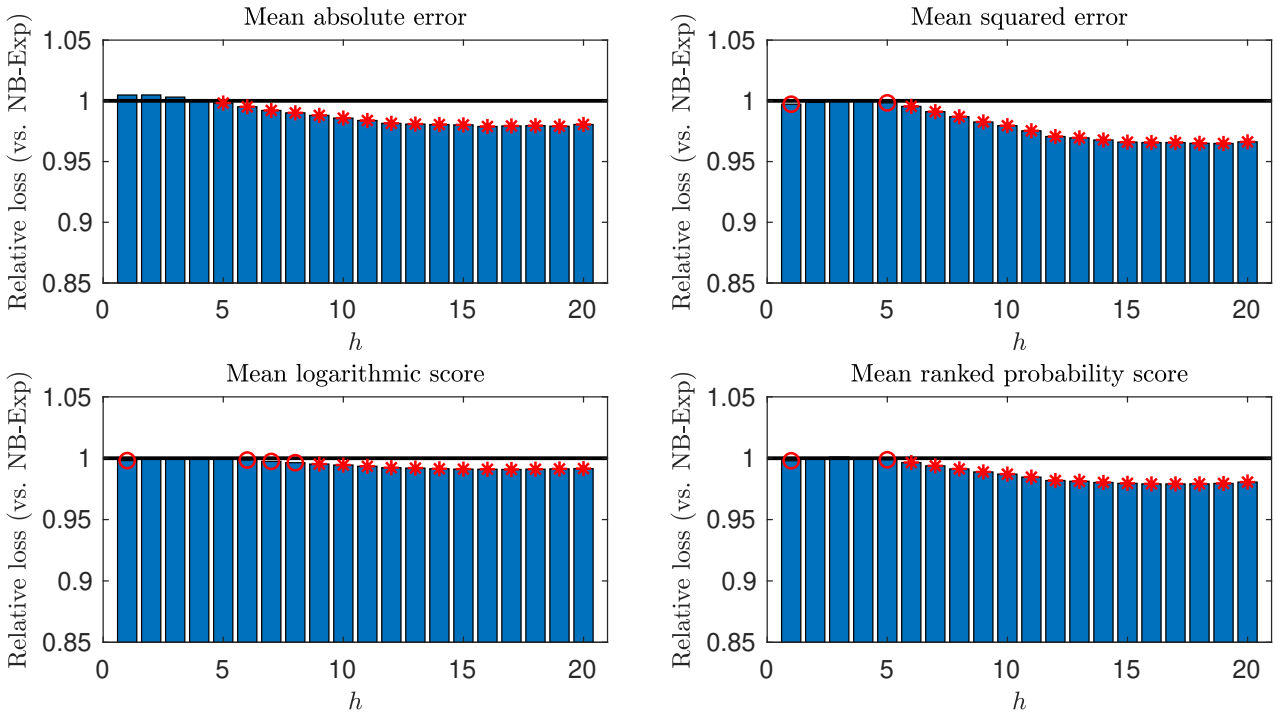

Figure S8: *Forecasting the spread level of the A stock on May 4, 2020. Four different loss metrics and twenty forecast horizons,* $h = 1, 2, \ldots, 20$. *The numbers plotted are relative average losses of the NB-Gamma forecasting model, compared with the NB-Exponential model, over* $n_{oos} = 720$ *out-of-sample forecasts. A circle above the bars indicates rejection null of equal forecasting performance between the two models, against the alternative that the NB-IG model provides superior forecasts, using the Diebold-Mariano (Diebold and Mariano, 1995) test at a* 5% *level; an asterisk denotes rejection at a* 1% *level.*

5 Details concerning integer-valued Lévy bases

5.1 Poisson Lévy basis

Consider the case where the Lévy basis is Poisson, i.e. $L' \sim \text{Poi}(\nu)$ for some intensity $\nu > 0$. For a bounded Borel set B with $Leb(B) < \infty$, we have

$$L(B) \sim \text{Poi}(\nu Leb(B)).$$

The cumulants in this case are $\kappa_j = \nu$ for all $j \geq 0$.

5.2 Negative binomial Lévy basis

We follow Barndorff-Nielsen et al. (2012, 2014) and denote by $NB(m, p)$ the negative binomial law with parameters $m \in \mathbb{N}$ and $p \in (0, 1)$. Recall, that a negative binomial random variable is positively valued and can be interpreted as the number of successes, k, until m failures in a sequence of iid Bernoulli trials, each with probability of success p. Let $L' \sim NB(m, p)$; it holds that

$$P(L' = k) = \frac{\Gamma(m+k)}{k!\Gamma(m)}(1-p)^m p^k, \quad k = 0, 1, 2, \ldots.$$

As is well known, we have that $L'_t \sim NB(mt, p)$ and therefore, for a Borel set B, it holds that $L(B) \sim NB(Leb(B)m, p)$, which implies

$$P(L(B) = k) = \frac{\Gamma(Leb(B)m+k)}{k!\Gamma(Leb(B)m)}(1-p)^{Leb(B)m} p^k, \quad k = 0, 1, 2, \ldots.$$

Here the relevant cumulants are $\kappa_1 = \frac{pm}{1-p}$, $\kappa_2 = \frac{pm}{(1-p)^2}$ and $\kappa_4 = m\frac{p+4p^2+p^3}{(1-p)^4}$.

5.3 Skellam Lévy basis

The Skellam distribution is the distribution of the difference of two Poisson processes N_t^+ and N_t^- and is therefore integer valued. Let $N_t^\pm \sim \text{Poi}(\psi^\pm)$ with $\psi^\pm > 0$; then $S := N_t^+ - N_t^- \sim \text{Skellam}(\psi^+, \psi^-)$. Further, the Skellam Lévy process $(L'_t)_{t\geq 0}$ with $L'_1 \sim \text{Skellam}(\psi^+, \psi^-)$ has the marginal distribution $L'_t \sim \text{Skellam}(t\psi^+, t\psi^-)$ (Barndorff-Nielsen et al., 2012), meaning that for a Borel set B, we have $L(B) \sim \text{Skellam}(Leb(B)\psi^+, Leb(B)\psi^-)$. The PMF of the random variable $X \sim \text{Skellam}(\psi^+, \psi^-)$ is given by

$$g(k; \psi^+, \psi^-) := P(X = k) = e^{-(\psi^+ + \psi^-)} \left(\frac{\psi^+}{\psi^-}\right)^{k/2} I_k\left(2\sqrt{\psi^+\psi^-}\right),$$

where $I_\nu(x)$ is the modified Bessel function of the first kind (see e.g. Abramowitz and Stegun (1972)) with parameter ν evaluated at x. In the symmetric case, $\psi^+ = \psi^- = \psi$, this reduces to $g(k; \psi) := e^{-2\psi} I_k(2\psi)$. The cumulants are easily seen to be $\kappa_j = \psi^+ - \psi^-$ for j odd and $\kappa_j = \psi^+ + \psi^-$ for j even.

5.4 ΔNB Lévy basis

Analogous to the Skellam process, we can consider the difference of two Lévy seeds which have negative binomials as their laws; Barndorff-Nielsen et al. (2012) call this a ΔNB Lévy process. Let $L^{\pm} \sim NB(m^{\pm}, p^{\pm})$ be independent Lévy seeds with negative binomial laws. Barndorff-Nielsen et al. (2012) show that for $k \geq 0$, the difference Lévy seed $L' = L^+ - L^-$ has PMF

$$P(L' = k) = (1-p^+)^{m^+}(1-p^-)^{m^-}\frac{(p^+)^k(m^+)_k}{k!}F(m^+ + k, m^-; k+1; p^+p^-), \tag{S1}$$

where

$$F(\alpha, \beta; \gamma; z) = \sum_{n=0}^{\infty}\frac{(\alpha)_n(\beta)_n}{(\gamma)_n}\frac{z^n}{n!}, \quad z \in [0,1), \quad \alpha, \beta, \gamma > 0,$$

is the hypergeometric function, see e.g. Abramowitz and Stegun (1972), and $(\alpha)_n = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}$ is the Pochhammer symbol. The PMF for $k \leq 0$ is, by symmetry, given as S1, mutatis mutandis. The resulting distribution is denoted as $L' \sim \Delta NB(m^+, p^+, m^-, p^-)$ and it is easy to show that (Barndorff-Nielsen et al., 2012) the Lévy process corresponding to L' has marginal distribution $L'_t \sim \Delta NB(tm^+, p^+, tm^-, p^-)$, meaning that we have for a Borel set B,

$$L(B) \sim NB(Leb(B)m^+, p^+, Leb(B)m^-, p^-).$$

The cumulants for the ΔNB Lévy seed is easily deduced from those of the negative binomial ones, recalling that the ΔNB law is the difference of two independent NB random variables.

6 Details concerning parametric trawl functions

The expressions for the likelihoods in the previous section reveal that we are interested in calculating expressions such as $Leb(A_t \backslash A)$ and $Leb(A_t \cap A)$ for different trawl functions. In this section we derive the required results for various trawls based on the superposition trawl function $d(s) = \int_0^\infty e^{\lambda s}\pi(d\lambda)$, $s \leq 0$, see also the main paper.

6.1 The exponential trawl

The case where the measure π has an atom at $\lambda > 0$, i.e. $\pi(dx) = \delta_\lambda(dx)$, where $\delta_y(\cdot)$ is the Dirac delta function at $y \in \mathbb{R}_+$, we get $d(s) = e^{\lambda s}$. Consequently, for $t \geq 0$,

$$Leb(A) = \lambda^{-1}, \quad Leb(A_t \backslash A) = \lambda^{-1}(1 - e^{-\lambda t}), \quad Leb(A_t \cap A) = \lambda^{-1}e^{-\lambda t}.$$

This implies the correlation function

$$\rho(h) = \exp(-\lambda h), \quad h > 0.$$

6.2 The finite superposition exponential trawl

Let π have finitely many atoms, i.e. $\pi(dx) = \sum_{i=1}^q w_i\delta_{\lambda_i}(dx)$ for $q \in \mathbb{N}$. Then

$$d(s) = \sum_{i=1}^q w_i e^{\lambda_i s},$$

and

$$Leb(A) = \sum_{i=1}^q w_i\lambda_i^{-1}, \quad Leb(A_t \backslash A) = \sum_{i=1}^q w_i\lambda_i^{-1}(1 - e^{-\lambda_i t}), \quad Leb(A_t \cap A) = \sum_{i=1}^q w_i\lambda_i^{-1}e^{-\lambda_i t}.$$

This implies the correlation function

$$\rho(h) = \left(\sum_{i=1}^q \omega_i\lambda_i^{-1}\right)^{-1} \sum_{i=1}^q \omega_i\lambda_i^{-1} \exp(-\lambda_i h), \quad h > 0.$$

6.3 The GIG trawl

A flexible class of trawl functions can be specified through the *generalized inverse Gaussian* (GIG) density function (see e.g. Barndorff-Nielsen et al. (2014)),

$$f_\pi(x) = \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)} x^{\nu-1} \exp\left(-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right),$$

where $\nu \in \mathbb{R}$ and $\gamma, \delta \geq 0$ with both not equal to zero simultaneously. $K_\nu(x)$ is the modified Bessel function of the third kind with parameter ν, evaluated at x (e.g. Abramowitz and Stegun (1972)). Suppose now, that π has density f_π, i.e. $\pi(d\lambda) = f_\pi(\lambda)d\lambda$. For $s \leq 0$, the trawl function becomes

$$d(s) = \int_0^\infty e^{\lambda s} f_\pi(\lambda)d\lambda = \left(1 - \frac{2s}{\gamma^2}\right)^{-\nu/2} \frac{K_\nu\left(\delta\gamma\alpha_s\right)}{K_\nu(\delta\gamma)},$$

whereas

$$Leb(A) = \frac{\gamma}{\delta}\frac{K_{\nu-1}(\delta\gamma)}{K_\nu(\delta\gamma)}, \quad Leb(A_t \cap A) = \frac{\gamma\alpha_t^{-\nu+1}}{\delta}\frac{K_{\nu-1}(\delta\gamma\alpha_t)}{K_\nu(\delta\gamma)},$$

and

$$Leb(A_t \backslash A) = \frac{\gamma}{\delta K_\nu(\delta\gamma)}\left(K_{\nu-1}(\delta\gamma) - \alpha_t^{-\nu+1}K_{\nu-1}(\delta\gamma\alpha_t)\right),$$

where $\alpha_t := \sqrt{\frac{2t}{\gamma^2}+1}$. This implies the correlation function

$$\rho(h) = \alpha_h^{-\nu+1}\frac{K_{\nu-1}(\delta\gamma\alpha_h)}{K_{\nu-1}(\delta\gamma)}, \quad h > 0.$$

6.4 The IG trawl

The inverse Gaussian distribution is a special case of the GIG distributions, where $\nu = \frac{1}{2}$. In this case, the trawl function simplifies to

$$d(s) = \left(1 - \frac{2s}{\gamma^2}\right)^{-1/2}\exp\left(\delta\gamma\left(1 - \sqrt{1 - \frac{2s}{\gamma^2}}\right)\right), \quad s \leq 0,$$

which means that

$$Leb(A) = \frac{\gamma}{\delta}, \quad Leb(A_t \cap A) = \frac{\gamma}{\delta}e^{\delta\gamma(1-\alpha_t)}, \quad Leb(A_t \backslash A) = \frac{\gamma}{\delta}\left(1 - e^{\delta\gamma(1-\alpha_t)}\right),$$

where again $\alpha_t = \sqrt{\frac{2t}{\gamma^2}+1}$. This implies the correlation function

$$\rho(h) = \exp(\delta\gamma(1-\alpha_h)), \quad h > 0.$$

6.5 The Γ trawl

An interesting case, capable of generating long memory in the trawl process, is given by the Γ trawl. Suppose that π has the $\Gamma(1+H, \alpha)$ density,

$$f_\pi(\lambda) = \frac{1}{\Gamma(1+H)}\alpha^{1+H}\lambda^H e^{-\lambda\alpha},$$

where $\alpha > 0$ and $H > 0$. Now,

$$d(s) = \left(1 - \frac{s}{\alpha}\right)^{-(H+1)}, \quad s \leq 0,$$

which implies

$$Leb(A) = \frac{\alpha}{H}, \quad Leb(A_t \cap A) = \frac{\alpha}{H}\left(1 + \frac{t}{\alpha}\right)^{-H}, \quad Leb(A_t \backslash A) = \frac{\alpha}{H}\left(1 - \left(1 + \frac{t}{\alpha}\right)^{-H}\right).$$

This yields the correlation function

$$\rho(h) = Corr(L(A_{t+h}), L(A_t)) = \frac{Leb(A_h \cap A)}{Leb(A)} = \left(1 + \frac{h}{\alpha}\right)^{-H},$$

so that

$$\int_0^\infty \rho(h)dh = \begin{cases} \infty & \text{if} \quad H \in (0,1], \\ \frac{\alpha}{H-1} & \text{if} \quad H > 1, \end{cases}$$

from which we see, that the trawl process has long memory for $H \in (0,1]$.

7 Details concerning gradients

Recall that we have the composite log likelihood function

$$l_{CL}(\theta; x) := l_{CL}^{(K)}(\theta; x) = \log L_{CL}^{(K)}(\theta; x) = \sum_{k=1}^{K} \sum_{i=1}^{n-k} \log f(x_{i+k}, x_i; \theta).$$

Let θ_i be an element of θ. The derivative of $l_{CL}(\theta; x)$ wrt. θ_i is

$$\frac{\partial}{\partial \theta_i} l_{CL}(\theta; x) = \frac{\partial}{\partial \theta_i} \log L_{CL}^{(K)}(\theta; x) = \sum_{k=1}^{K} \sum_{i=1}^{n-k} \frac{1}{f(x_{i+k}, x_i; \theta)} \frac{\partial}{\partial \theta_i} f(x_{i+k}, x_i; \theta). \tag{S1}$$

Recall also that

$$f(x_{i+k}, x_i; \theta) = \sum_{c=-\infty}^{\infty} P_{1,i,k}^{(c)} \cdot P_{2,i,k}^{(c)} \cdot P_{3,k}^{(c)}$$

with

$$P_{1,i,k}^{(c)} := \mathbb{P}\left(L(A_{k\Delta} \setminus A) = x_{i+k} - c\right), \quad P_{2,i,k}^{(c)} = \mathbb{P}\left(L(A_{k\Delta} \setminus A) = x_i - c\right), \quad P_{3,k}^{(c)} = \mathbb{P}\left(L(A_{k\Delta} \cap A) = c\right),$$

implying that

$$\frac{\partial}{\partial \theta_i} f(x_{i+k}, x_i; \theta) = \sum_{c=-\infty}^{\infty} \left(\frac{\partial}{\partial \theta_i} P_{1,i,k}^{(c)} \cdot P_{2,i,k}^{(c)} \cdot P_{3,k}^{(c)} + P_{1,i,k}^{(c)} \cdot \frac{\partial}{\partial \theta_i} P_{2,i,k}^{(c)} \cdot P_{3,k}^{(c)} + P_{1,i,k}^{(c)} \cdot P_{2,i,k}^{(c)} \cdot \frac{\partial}{\partial \theta_i} P_{3,k}^{(c)} \right). \tag{S2}$$

The terms $P_{1,i,k}^{(c)}, P_{2,i,k}^{(c)}, P_{3,k}^{(c)}$ are calculated in the numerical maximization of the composite likelihood routine for all c. The aim of this section is to calculate $\frac{\partial}{\partial \theta_i} P_{j,i,k}^{(c)}$ for $j = 1, 2, 3$, so that the gradient of the log likelihood function is easily calculated using Equations (S1) and (S2). It is clear that $\frac{\partial}{\partial \theta_i} P_{j,i,k}^{(c)}$ will depend on both the Lévy basis as well as the form of the trawl set (and hence the trawl function). We first supply the relevant derivations for the Poisson Lévy basis (Section 7.2) and the Negative Binomial Lévy basis (Section 7.3), and then the trawl functions Exp, SupExp, IG, and Γ (Sections 7.4–7.7).

7.1 Some preliminary practical details

In our numerical implementation of the composite likelihood methods, we often have restrictions on some parameters. Most notably, we have positivity restriction, e.g. we require that the intensity $\nu > 0$ for the Poisson Lévy basis. One could impose such restrictions buy using a constrained optimization procedure, when performing the numerical optimization of the log composite likelihood function $l_{CL}(\theta; x)$. We prefer to work with an unconstrained optimization procedure, by transforming the parameters such that they are fulfilling their restrictions automatically. That is, if θ is a restricted parameter, we find an invertible transformation function g, such that $\tilde{\theta} = g^{-1}(\theta) \in \mathbb{R}$ is

unrestricted. The unconstrained numerical optimizer is optimizing over the unrestricted parameter $\tilde{\theta}$ and arrives at, say, $\tilde{\theta}^*$. Our estimate of θ is thus $\hat{\theta} = g(\tilde{\theta}^*)$. Consequently, it is necessary to correct for this when calculating standard errors (delta rule) as well as when supplying a gradient for our numerical optimization scheme. The reason is that the calculations concerning the gradient, detailed in the previous section, are with respect to θ, and not $\tilde{\theta}$, which is the actual parameter being used in the numerical optimization procedure. In case of a transformed variable, the gradient that should be supplied to the machine is therefore not the one given in (S2), but rather

$$\frac{\partial}{\partial \tilde{\theta}_i} f(x_{i+k}, x_i; \theta) = \frac{\partial}{\partial \theta_i} f(x_{i+k}, x_i; \theta) \frac{\partial \theta_i}{\partial \tilde{\theta}_i} = \frac{\partial}{\partial \theta_i} f(x_{i+k}, x_i; \theta) \frac{\partial}{\partial \tilde{\theta}_i} g(\tilde{\theta}).$$

In this paper two restrictions are encountered: many parameters are positive, while a few are restricted to be in the unit interval. If $\theta > 0$ is a positive parameter, we use a log transformation by defining the new parameter $\tilde{\theta}$ through

$$\tilde{\theta} = g^{-1}(\theta) = \log \theta, \qquad \theta = g(\tilde{\theta}) = \exp(\tilde{\theta}).$$

If $p \in (0,1)$ is a parameter, we use an inverse logistic (sigmoid) transformation,

$$\tilde{p} = g^{-1}(p) = \log\left(\frac{p}{1-p}\right), \qquad p = g(\tilde{p}) = \frac{1}{1+\exp(-\tilde{p})}.$$

7.2 Poisson Lévy basis

Let $L' \sim Poi(\nu)$ and recall that for a Borel set B, this implies

$$\mathbb{P}\left(L(B) = x\right) = \frac{[\nu Leb(B)]^x \exp(-\nu Leb(B))}{x!}.$$

We deduce, that for a generic parameter $\theta \neq \nu$,

$$\frac{\partial}{\partial \theta} \mathbb{P}\left(L(B) = x\right) = \left(x Leb(B)^{-1} - \nu\right) \mathbb{P}\left(L(B) = x\right) \frac{\partial}{\partial \theta} Leb(B).$$

The only ingredient left to calculate is

$$\frac{\partial}{\partial \nu} \mathbb{P}\left(L(B) = x\right) = \left(x\nu^{-1} - Leb(B)\right) \mathbb{P}\left(L(B) = x\right).$$

7.3 Negative Binomial Lévy basis

Recall that in the case where the Lévy seed L' is distributed as a Negative Binomial random variable with parameters $m > 0$ and $p \in [0,1]$, we have $L(B) \sim NB(Leb(B)m, p)$, which implies

$$P(L(B) = x) = \frac{\Gamma(Leb(B)m + x)}{x!\Gamma(Leb(B)m)}(1-p)^{Leb(B)m} p^x, \quad x = 0, 1, 2, \ldots.$$

Using the well-known property of the Γ function that $\Gamma(x+1) = x\Gamma(x)$ (Gradshteyn and Ryzhik, 2007, p. 904), we can write

$$P(L(B) = x) = \left(Leb(B)m + x - 1\right)\left(Leb(B)m + x - 2\right) \cdots \left(Leb(B)m\right) \frac{1}{x!}(1-p)^{Leb(B)m} p^x,$$

for $k = 0, 1, 2, \ldots$. We deduce, that for a generic parameter $\theta \neq m, p$,

$$\frac{\partial}{\partial\theta}P(L(B) = x)$$
$$= \left(\frac{\partial}{\partial\theta}Leb(B)\right) mP(L(B) = x)\left(\log(1-p) + \frac{1}{Leb(B)m} + \frac{1}{Leb(B)m+1} + \cdots + \frac{1}{Leb(B)m+x-1}\right).$$

The only ingredients left to calculate are

$$\frac{\partial}{\partial p}P(L(B) = x) = P(L(B) = x)\left(\frac{x}{p} - \frac{Leb(B)m}{1-p}\right),$$

and

$$\frac{\partial}{\partial m}P(L(B) = x)$$
$$= P(L(B) = x)Leb(B)\left(\log(1-p) + \frac{1}{Leb(B)m} + \frac{1}{Leb(B)m+1} + \cdots + \frac{1}{Leb(B)m+x-1}\right).$$

7.4 Exponential trawl function

Let L' be a generic Lévy seed and $d(s) = \exp(\lambda s)$ for $s \leq 0$. Recall that for $t > 0$,

$$Leb(A_t \setminus A) = \lambda^{-1}(1 - \exp(-\lambda t)), \qquad Leb(A_t \cap A) = \lambda^{-1}\exp(-\lambda t).$$

It is not difficult to show that

$$\frac{\partial}{\partial\lambda}Leb(A_t \setminus A) = \lambda^{-1}\left(t\exp(-\lambda t) - \lambda^{-1}(1 - \exp(-\lambda t))\right),$$

while

$$\frac{\partial}{\partial\lambda}Leb(A_t \cap A) = -\lambda^{-1}\exp(-\lambda t)\left(\lambda^{-1} + t\right).$$

7.5 SupExp trawl function

Let L' be a generic Lévy seed and $d(s)$ be the supExp trawl function (see above). Recall that for $t > 0$,

$$Leb(A_t \setminus A) = \sum_{i=1}^{q} w_i\lambda_i^{-1}(1 - e^{-\lambda_i t}), \qquad Leb(A_t \cap A) = \sum_{i=1}^{q} w_i\lambda_i^{-1}e^{-\lambda_i t}.$$

It is not difficult to show that for $j = 1, 2, \ldots, q$,

$$\frac{\partial}{\partial\lambda_j}Leb(A_t \setminus A) = w_j\lambda_j^{-1}\left(t\exp(-\lambda_j t) - \lambda_j^{-1}(1 - \exp(-\lambda_j t))\right),$$
$$\frac{\partial}{\partial\lambda_j}Leb(A_t \cap A) = -w_j\lambda_j^{-1}\exp(-\lambda_j t)\left(\lambda_j^{-1} + t\right),$$

while

$$\frac{\partial}{\partial w_j}Leb(A_t \setminus A) = \lambda_j^{-1}(1 - e^{-\lambda_j t}),$$
$$\frac{\partial}{\partial w_j}Leb(A_t \cap A) = \lambda_j^{-1}e^{-\lambda_j t}.$$

7.6 IG trawl function

Let L' be a generic Lévy seed and $d(s)$ be the IG trawl (see above). Recall that, for $t > 0$,

$$Leb(A_t \setminus A) = \frac{\gamma}{\delta}\left(1 - \exp(\delta\gamma(1-\alpha_t))\right), \quad Leb(A_t \cap A) = \frac{\gamma}{\delta}\exp(\delta\gamma(1-\alpha_t)),$$

where $\alpha_t = \sqrt{\frac{2t}{\gamma^2} + 1}$.

We can show that

$$\frac{\partial}{\partial\delta}Leb(A_t \setminus A) = -\delta^{-1}Leb(A_t \setminus A) - \gamma^2\delta^{-1}(1-\alpha_t)\exp(\delta\gamma(1-\alpha_t)),$$
$$\frac{\partial}{\partial\gamma}Leb(A_t \setminus A) = -\gamma^{-1}Leb(A_t \setminus A) - \gamma\exp(\delta\gamma(1-\alpha_t))[1 - \alpha_t + 2\gamma^{-2}\alpha_t^{-1}t],$$

and

$$\frac{\partial}{\partial\delta}Leb(A_t \cap A) = Leb(A_t \cap A)(\gamma(1-\alpha_t) - \delta^{-1}),$$
$$\frac{\partial}{\partial\gamma}Leb(A_t \cap A) = Leb(A_t \cap A)(\gamma^{-1} + \delta(1-\alpha_t) + 2\delta\gamma^{-2}\alpha_t^{-1}t).$$

7.7 Γ trawl function

Let L' be a generic Lévy seed and $d(s)$ be the Γ trawl (see above). Recall that, for $t > 0$,

$$Leb(A_t \setminus A) = \frac{\alpha}{H}\left(1 - \left(1 + \frac{t}{\alpha}\right)^{-H}\right), \quad Leb(A_t \cap A) = \frac{\alpha}{H}\left(1 + \frac{t}{\alpha}\right)^{-H}.$$

It is easy to show that

$$\frac{\partial}{\partial H}Leb(A_t \cap A) = -\alpha\left(1 + \frac{t}{\alpha}\right)^{-(H+1)}\left(H^{-2}\left(1 + \frac{t}{\alpha}\right) + 1\right),$$
$$\frac{\partial}{\partial\alpha}Leb(A_t \cap A) = \left(1 + \frac{t}{\alpha}\right)^{-(H+1)}\left(H^{-1}\left(1 + \frac{t}{\alpha}\right) + \frac{t}{\alpha}\right).$$

and

$$\frac{\partial}{\partial H}Leb(A_t \setminus A) = -\alpha H^{-2} - \frac{\partial}{\partial H}Leb(A_t \cap A),$$
$$\frac{\partial}{\partial\alpha}Leb(A_t \setminus A) = H^{-1} - \frac{\partial}{\partial\alpha}Leb(A_t \cap A).$$

8 Additional calculations

8.1 Calculations for the GIG trawl of Section 6.3

We have the trawl function

$$d(s) = \left(1 - \frac{2s}{\gamma^2}\right)^{-\nu/2} \frac{K_\nu\left(\delta\gamma\sqrt{1-\frac{2s}{\gamma^2}}\right)}{K_\nu(\delta\gamma)}.$$

In the following we use the substitution $x = \sqrt{1+\frac{2s}{\gamma^2}}$ to get

$$\begin{aligned} Leb(A) = \int_0^\infty d(-s)ds &= \int_0^\infty \left(1+\frac{2s}{\gamma^2}\right)^{-\nu/2} \frac{K_\nu\left(\delta\gamma\sqrt{1+\frac{2s}{\gamma^2}}\right)}{K_\nu(\delta\gamma)} ds \\ &= \int_1^\infty x^{-\nu+1} \frac{K_\nu(\delta\gamma x)}{K_\nu(\delta\gamma)} \gamma^2 dx \\ &= \frac{\gamma^2}{K_\nu(\delta\gamma)} \left(\int_0^\infty x^{-\nu+1} K_\nu(\delta\gamma x) - \int_0^1 x^{-\nu+1} K_\nu(\delta\gamma x) \right). \end{aligned}$$

Now apply (6.561.12) and (6.561.16) in Gradshteyn and Ryzhik (2007) to get[15]

$$Leb(A) = \frac{\gamma}{\delta} \frac{K_{\nu-1}(\delta\gamma)}{K_\nu(\delta\gamma)}.$$

Set $\alpha := \sqrt{\frac{2t}{\gamma^2}+1}$. Now, make the same substitution as above to get

$$\begin{aligned} Leb(A_t \cap A) = \int_t^\infty d(-s)ds &= \int_0^\infty \left(1+\frac{2s}{\gamma^2}\right)^{-\nu/2} \frac{K_\nu\left(\delta\gamma\sqrt{1+\frac{2s}{\gamma^2}}\right)}{K_\nu(\delta\gamma)} ds \\ &= \int_\alpha^\infty x^{-\nu+1} \frac{K_\nu(\delta\gamma x)}{K_\nu(\delta\gamma)} \gamma^2 dx. \end{aligned}$$

Set $y = \alpha^{-1}x$ to get

$$\begin{aligned} \int_\alpha^\infty x^{-\nu+1} \frac{K_\nu(\delta\gamma x)}{K_\nu(\delta\gamma)} \gamma^2 dx &= \frac{\gamma^2}{K_\nu(\delta\gamma)} \int_1^\infty (\alpha y)^{-\nu+1} K_\nu(\delta\gamma\alpha y)\, \alpha dy \\ &= \frac{\gamma^2\alpha^{-\nu+2}}{K_\nu(\delta\gamma)} \int_1^\infty y^{-\nu+1} K_\nu(\delta\gamma\alpha y)\, dy. \end{aligned}$$

Now, splitting the integral as above and using the same formulae yields

$$Leb(A_t \cap A) = \frac{\gamma\alpha^{-\nu+1}}{\delta} \frac{K_{\nu-1}(\delta\gamma\alpha)}{K_\nu(\delta\gamma)}.$$

8.2 Calculations for the IG trawl of Section 6.4

We have

$$d(s) = \left(1 - \frac{2s}{\gamma^2}\right)^{-1/2} \exp\left(\delta\gamma\left(1 - \sqrt{1-\frac{2s}{\gamma^2}}\right)\right),$$

[15] Note, that we here need to impose $\nu < 1$.

which means that

$$Leb(A) = \int_0^\infty d(-s)ds = \int_0^\infty \left(1 + \frac{2s}{\gamma^2}\right)^{-1/2} \exp\left(\delta\gamma\left(1 - \sqrt{1 + \frac{2s}{\gamma^2}}\right)\right) ds.$$

So, after the change of variable $x = \sqrt{1 + \frac{2s}{\gamma^2}}$ we have

$$\begin{aligned} Leb(A) = \int_0^\infty d(-s)ds &= \int_1^\infty x^{-1} \exp\left(\delta\gamma(1-x)\right) \gamma^2 x dx \\ &= \gamma^2 \int_1^\infty \exp\left(\delta\gamma(1-x)\right) dx \\ &= \gamma^2 e^{\delta\gamma} \int_1^\infty e^{-\delta\gamma x} dx \\ &= \frac{\gamma}{\delta}. \end{aligned}$$

Again, defining $\alpha := \sqrt{\frac{2t}{\gamma^2} + 1}$, we get by similar calculations

$$Leb(A_t \cap A) = \int_t^\infty d(-s)ds = \gamma^2 e^{\delta\gamma} \int_\alpha^\infty e^{-\delta\gamma x} dx = \frac{\gamma}{\delta} e^{\delta\gamma(1-\alpha)}.$$

9 Software (MATLAB)

The following functions are available in the MATLAB software language. We give a very brief description of the functions here, but refer to the extensive documentation in the code for further details. The code can be freely downloaded from the webpage https://github.com/mbennedsen/Likelihood-based-IVT.

- `simulateIVT`:
  - Simulates equidistant observations of a parametric IVT process, specified by a Lévy basis and a trawl function. The Lévy basis and trawl function can be specified independently of each other using the framework described in this Supplementary Material.
- `estimateIVT`:
  - Takes as input a vector of equidistantly spaced observations and a parametric specification (Lévy basis and trawl function) and outputs estimates of the corresponding parameters using the maximum composite likelihood approach developed in the main paper.
- `modelselectIVT`:
  - This function estimates six parametric IVT models (Poisson-Exponential, Poisson-IG, Poisson-Gamma, NB-Exponential, NB-IG, NB-Gamma) and calculates the composite likelihood function when evaluated in the optimized parameters, as well as the CLAIC and CLBIC criteria given in the main paper. These three criteria can be used for model selection, larger values indicating a better fit.
- `forecastIVT`:
  - Takes as input a parametric IVT model (Lévy basis and trawl function), a forecast horizon (which can be a vector of several forecast horizons), as well as historical observations; the output is the predictive probability distribution for the given forecast horizons. The parameters underlying the predictive distribution are estimated using the maximum composite likelihood approach presented in the main paper.
- `analyze_stock_A` and `analyze_simulated_data`:
  - These files illustrate the use of the functions `simulateIVT`, `estimateIVT`, `modelselectIVT`, and `forecastIVT`. The file `analyze_stock_A` reproduces the output of the main paper, while `analyze_simulated_data` simulates a user-specified IVT process and then conducts analyses similar to those considered in the main paper on these simulated data.

References

Abramowitz, M. and I. A. Stegun (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables* (10th ed.), Volume 55. United States Department of Commerce.

Barndorff-Nielsen, O. E., A. Lunde, N. Shephard, and A. Veraart (2014). Integer-valued trawl processes: A class of stationary infinitely divisible processes. *Scandinavian Journal of Statistics 41*, 693–724.

Barndorff-Nielsen, O. E., D. G. Pollard, and N. Shephard (2012). Integer-valued Lévy processes and low latency financial econometrics. *Quantitative Finance 12*, 587–605.

Bennedsen, M., A. Lunde, N. Shephard, and A. Veraart (2021). Inference and forecasting for continuous-time integer-valued trawl processes and their use in financial economics. Preprint.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1989). *Regular Variation*. Cambridge University Press.

Davidson, J. (1994). *Stochastic Limit Theory: Introduction for Econometricians*. Advanced Texts in Econometrics. Oxford University Press.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Gradshteyn, I. S. and I. M. Ryzhik (2007). *Table of integrals, series, and products* (Seventh ed.). Amsterdam: Academic Press.

Newey, W. K. and D. McFadden (1994). Chapter 36: Large sample estimation and hypothesis testing. Volume 4 of *Handbook of Econometrics*, pp. 2111 – 2245. Elsevier.

Wooldridge, J. M. (1994). Chapter 45: Estimation and inference for dependent processes. Volume 4 of *Handbook of Econometrics*, pp. 2639 – 2738. Elsevier.

Research Papers 2021

CREATES
Center for Research in Econometric Analysis of Time Series

2020-16: J. Eduardo Vera-Valdés: Temperature Anomalies, Long Memory, and Aggregation

2020-17: Jesús-Adrián Álvarez, Malene Kallestrup-Lamb and Søren Kjærgaard: Linking retirement age to life expectancy does not lessen the demographic implications of unequal lifespans

2020-18: Mikkel Bennedsen, Eric Hillebrand and Siem Jan Koopman: A statistical model of the global carbon budget

2020-19: Eric Hillebrand, Jakob Mikkelsen, Lars Spreng and Giovanni Urga: Exchange Rates and Macroeconomic Fundamentals: Evidence of Instabilities from Time-Varying Factor Loadings

2021-01: Martin M. Andreasen: The New Keynesian Model and Bond Yields

2021-02: Daniel Borup, David E. Rapach and Erik Christian Montes Schütte: Now- and Backcasting Initial Claims with High-Dimensional Daily Internet Search-Volume Data

2021-03: Kim Christensen, Mathias Siggaard and Bezirgen Veliyev: A machine learning approach to volatility forecasting

2021-04: Fabrizio Iacone, Morten Ørregaard Nielsen and Robert Taylor: Semiparametric Tests for the Order of Integration in the Possible Presence of Level Breaks

2021-05: Stefano Grassi and Francesco Violante: Asset Pricing Using Block-Cholesky GARCH and Time-Varying Betas

2021-06: Gloria González-Rivera, Carlos Vladimir Rodríguez-Caballero and Esther Ruiz Ortega: Expecting the unexpected: economic growth under stress

2021-07: Matei Demetrescu and Robinson Kruse-Becher: Is U.S. real output growth really non-normal? Testing distributional assumptions in time-varying location-scale models

2021-08: Luisa Corrado, Stefano Grassi and Aldo Paolillo: Modelling and Estimating Large Macroeconomic Shocks During the Pandemic

2021-09: Leopoldo Catania, Alessandra Luati and Pierluigi Vallarino: Economic vulnerability is state dependent

2021-10: Søren Johansen and Anders Ryghn Swensen: Adjustment coefficients and exact rational expectations in cointegrated vector autoregressive models

2021-11: Bent Jesper Christensen, Mads Markvart Kjær and Bezirgen Veliyev: The incremental information in the yield curve about future interest rate risk

2021-12: Mikkel Bennedsen, Asger Lunde, Neil Shephard and Almut E. D. Veraart: Inference and forecasting for continuous-time integer-valued trawl processes and their use in financial economics