

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


Integer-valued trawl processes: A class of stationary infinitely divisible processes

Citation

Barndorff-Nielsen, Ole E., Asger Lunde, Neil Shephard, and Almut E.D. Veraart. 2014. Integer-valued trawl processes: A class of stationary infinitely divisible processes. Scandinavian Journal of Statistics 41 (3): 693–724. doi:10.1111/sjos.12056.

Published version

https://doi.org/10.1111/sjos.12056

Link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:34650304

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles (OAP), as set forth at https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story

Integer-valued trawl processes: A class of stationary infinitely divisible processes

Ole E. Barndorff-Nielsen
The T.N. Thiele Centre for Mathematics in Natural Science,
Department of Mathematical Sciences, Aarhus University,
Ny Munkegade, DK-8000 Aarhus C, Denmark
& CREATES, Aarhus University
oebn@imf.au.dk

Asger Lunde
CREATES, Department of Economics and Business, Aarhus University,
Fulgesangs Allé 4, DK-8210 Aarhus V, Denmark
alunde@econ.au.dk

Neil Shephard
Department of Economics, Harvard University, Littauer Center
and Department of Statistics, Harvard University, Science Center
Cambridge, MA 02138, USA
shephard@fas.harvard.edu

Almut E. D. Veraart
Department of Mathematics, Imperial College London, & CREATES
180 Queen's Gate, SW7 2AZ London, UK
a.veraart@imperial.ac.uk

September 9, 2013

Abstract

This paper introduces a new continuous-time framework for modelling serially correlated count and integer-valued data. The key component in our new model is the class of integer-valued trawl (IVT) processes, which are serially correlated, stationary, infinitely divisible processes. We analyse the probabilistic properties of such processes in detail and, in addition, study volatility modulation and multivariate extensions within the new modelling framework. Moreover, we describe how the parameters of a trawl process can be estimated and obtain promising estimation results in our simulation study. Finally, we apply our new modelling framework to high frequency financial data.

Keywords: Lévy bases; Trawl processes; Stationarity; Stochastic volatility, Time change.

1 Introduction

This paper introduces a new way of building a continuous-time framework for serially correlated count data. Count data appear in various applications including medical science, epidemiology, meteorology, network modelling, actuarial science, econometrics and finance.

There has been very active research on various aspects of count data, see e.g. Cameron & Trivedi (1998), Kedem & Fokianos (2002) for textbook treatments and Cui & Lund (2009),

Davis et al. (1999), Davis & Wu (2009), Jung & Tremayne (2011), McKenzie (2003), Weiß (2008) for recent surveys and some new developments of the literature. Count data are characterised as being non-negative, integer-valued, and often over-dispersed, meaning that the variance is typically greater than the corresponding mean.

In this paper, we focus on count data which exhibit a serial dependence, i.e. time-series data. Due to the great variety of applications, the literature on count data is spread across different disciplines. Overall, we find that two modelling approaches seem to be predominant in the literature: First, there is the class of so-called discrete autoregressive moving-average (DARMA) models introduced by Jacobs & Lewis (1978*a*,*b*). McKenzie (2003) pointed out that the advantage of such stationary processes is that their marginal distribution can be of any kind. However, this comes at the cost that the dependence structure is generated by potentially long runs of constant values, which results in sample paths which are rather unrealistic in many applications.

Second, there is the wide class of models obtained from thinning operations. These models go back to the influential work of Steutel & van Harn (1979), who developed discrete analogues of self-decomposability and stability. See also Zhu & Joe (2003) for related more recent work. Thinning operations include e.g. binomial thinning, hypergeometric thinning, random coefficient thinning, iterated thinning and quasi-binomial thinning, see Weiß (2008) for a detailed review. Sample paths of thinned models often appear to be more realistic than their counterparts stemming from DARMA processes. However, thinned models cannot generate an arbitrary marginal distribution for integer-valued data.

Further models for count data include regression type models, which are typically based on generalised linear models. Also, state-space and Bayesian approaches have recently been developed in the literature, see McKenzie (2003) for a review of these models.

The new class of stationary integer-valued trawl (IVT) processes which we develop in this paper allows for a very flexible autocorrelation structure and at the same time can generate any kind of marginal distribution within the class of integer-valued infinitely divisible distributions. IVT processes are nested within the wide class of so-called ambit fields, see e.g. Barndorff-Nielsen et al. (2011), Barndorff-Nielsen & Schmiegel (2007).

A key aspect which makes this work distinct from other papers on integer-valued processes is the fact that we can account for stochastic volatility. Stochastic volatility (or intermittency) has been studied extensively in the last decade both in the finance literature as well as in physics in the context of modelling turbulence, see e.g. Barndorff-Nielsen & Shephard (2013) for a textbook treatment in the finance context and see Barndorff-Nielsen, Benth & Veraart (2012) for relevant references in the turbulence context. In this paper, we discuss various new methods for accounting for stochastic volatility clusters in the context of integer-valued stochastic processes.

Moreover, we present extensions to a multivariate modelling framework, and we carry out a simulation study, where we demonstrate how our new class of models can be estimated in practice. In an empirical study of high frequency financial data we illustrate the relevance of our new class of processes for empirical work.

The outline for the remaining part of the paper is as follows. Section 2 introduces the new class of integer-valued trawl (IVT) processes. We will see that an IVT process is specified by its marginal infinitely divisible distribution and by its trawl. Hence Section 3 focuses on relevant classes of infinitely divisible distributions suitable for modelling count data, and Section 4 presents suitable choices of the trawl. Next, Section 5 shows how stochastic volatility can be introduced and Section 6 presents multivariate extensions of the trawl framework. We carry out a simulation study in Section 7, where we demonstrate how IVT processes can be estimated in practice. Moreover, an empirical illustration in the context of high frequency financial data is given in Section 8, and, finally, Section 9 concludes. Proofs and detailed computations for

various examples in the paper are relegated to Section A in the Appendix.

2 Integer-valued trawl processes

While general trawl processes have recently been introduced in the literature, see Barndorff-Nielsen (2011) and also Barndorff-Nielsen, Benth & Veraart (2012) for detailed background material, the focus of this paper is on the important sub-class of *integer-valued trawl* (IVT) processes. Such processes are stationary, integer-valued stochastic processes, which allow for serial correlation and are infinitely divisible.

2.1 Integer-valued, homogeneous Lévy bases

Throughout the paper let $(\Omega, \mathcal{F}, \mathbb{P})$ denote the probability space. Further, let $E \subset \mathbb{R}^d$ $(d \in \mathbb{N})$ and let the corresponding Borel σ-algebra be denoted by $\mathcal{E} = \mathcal{B}(E)$. Further μ denotes a Radon measure on $(E, \mathcal{E})$, i.e. $\mu(B) < \infty$ for every compact measurable set $B \in \mathcal{E}$.

We start by recalling the definition of a Poisson random measure, cf. eg. Cont & Tankov (2004, p. 57).

Definition 1 *A* Poisson random measure *on E with intensity measure μ is defined as an integer-valued random measure $N : \Omega \times \mathcal{E} \to \mathbb{N}_0$, $(\omega, B) \mapsto N(\omega, B)$ satisfying the following three conditions:*

1. *For any bounded measurable set $B \subset E$, $N(B) < \infty$ is an integer-valued random variable; and for (almost all) $\omega \in \Omega$, $N(\omega, \cdot)$ is an integer-valued Radon measure on E.*
2. *For each measurable set $B \subset E$, the random variable $N(\cdot, B) = N(B)$ is Poisson distributed with parameter $\mu(B)$.*
3. *For any mutually disjoint measurable sets $B_1, \ldots, B_n \in \mathcal{E}$, $n \in \mathbb{N}$, the random variables $N(B_1), \ldots, N(B_n)$ are independent.*

In the following, we are interested in certain types of Poisson random measures. For that, let us recall the definition of a Lévy measure.

Definition 2 *Let ν denote a Radon measure on $\mathbb{R}$ satisfying $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}} \min(|x|^2, 1)\nu(dx) < \infty$. Then ν is called a* Lévy measure.

Throughout this paper, we will be working under the following assumption.

Assumption (A) Let N be a (homogeneous) Poisson random measure on $E = \mathbb{R} \times [0, 1] \times \mathbb{R}$ with intensity measure $\mu(dy, dx, ds) = \mathbb{E}(N(dy, dx, ds)) = \nu(dy)dxds$, where ν is a Lévy measure concentrated on $\mathbb{Z} \setminus \{0\}$ and satisfying $\int_{-\infty}^{\infty} \min(1, |y|)\nu(dy) < \infty$.

The key component for an IVT process is an integer-valued, homogeneous Lévy basis, which we will define in terms of a Poisson random measure.

Definition 3 *Let N be a homogeneous Poisson random measure on $(E, \mathcal{E})$ satisfying Assumption (A). Then we define an* integer-valued, homogeneous Lévy basis *on $([0, 1] \times \mathbb{R}, \mathcal{B}([0, 1] \times \mathbb{R}))$ in terms of the Poisson random measure as*

$$L(dx, ds) = \int_{-\infty}^{\infty} yN(dy, dx, ds). \tag{1}$$

Clearly, L is infinitely divisible (ID)[1] with characteristic function $\mathbb{E}(\exp(i\theta L(dx,ds))) = \exp(C(\theta \ddagger L(dx,ds)))$, where

$$C(\theta \ddagger L(dx,ds)) = \int_{\mathbb{R}} \left(e^{i\theta y} - 1\right) \nu(dy)dxds \tag{2}$$

denotes the corresponding cumulant function, i.e. the distinguished logarithm of the characteristic function, cf. Sato (1999, p. 33).

Definition 4 *Let L be an integer-valued, homogeneous Lévy basis with cumulant function given by (2). Then the random variable L' satisfying $\mathbb{E}(\exp(i\theta L')) = \exp(C(\theta \ddagger L'))$, with $C(\theta \ddagger L') = \int_{\mathbb{R}} \left(e^{i\theta y} - 1\right) \nu(dy)$, is called the* Lévy seed *associated with the Lévy basis L.*

Then we can rewrite equation (2) as

$$C(\theta \ddagger L(dx,ds)) = C(\theta \ddagger L')dxds. \tag{3}$$

From the expressions for the cumulant functions in (3) we see that the Lévy seed represents the infinitely divisible random variable which specifies the law of the corresponding Lévy basis. Likewise, we find that the law of the homogeneous Lévy basis L completely determines the law of the corresponding Lévy seed L'. We will use this important property throughout the paper. In particular, we will typically specify a model in terms of the Lévy seed L' and then assume that the law of the corresponding Lévy basis is determined by (3).

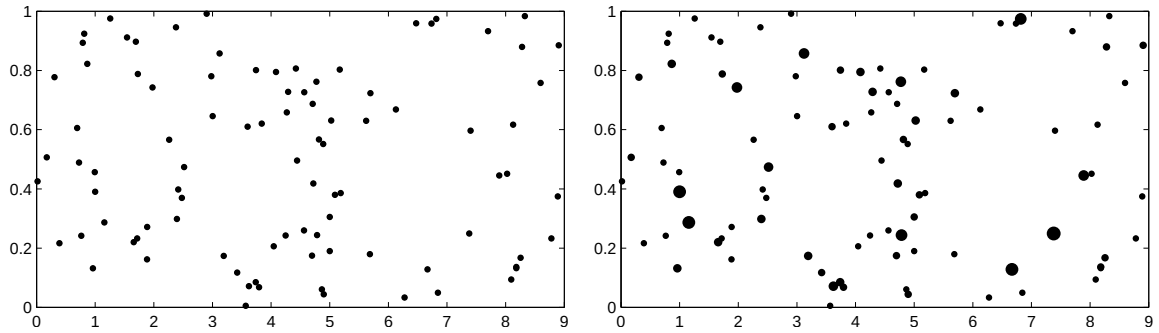
Remark 1 *We can associate a Lévy process denoted by $(L'_t)_{t\geq 0}$ with the Lévy seed L' by defining $L'_1 = L'$. Then $C(\theta \ddagger L'_t) = tC(\theta \ddagger L')$. So the Lévy seed can simply be understood as the ID random variable which determines an ID law from which we can construct either a Lévy process or, more generally, a Lévy basis. In that context, the corresponding Lévy measure $\nu(B)$ represents the expected number of jumps, in a unit time interval, whose size belongs to B, where $B \in \mathcal{B}(\mathbb{R})$ is a Borel set.*

Note that a given realisation of the Lévy basis determines a countable set R of points (y,x,s) in $\mathbb{Z}\setminus\{0\} \times [0,1] \times \mathbb{R}$, where the last two coordinates (x,s) are uniformly distributed over $[0,1]\times\mathbb{R}$ while y is the value of the basis at that point. The projection of R onto the time axis constitutes the realisation of a (two-sided) Poisson process with intensity $\int_{-\infty}^{\infty} \nu(dy)$. Now let (y,x,s) be a point in R. In the case of a Poisson Lévy basis, we always have $y = 1$, whereas for general integer-valued Lévy bases any $y \in \mathbb{Z}\setminus\{0\}$ can potentially occur. Figure 1 illustrates this feature. The picture on the left hand side shows a realisation of a Poisson Lévy basis, whereas the graph on the right hand side shows a realisation of a negative Binomial Lévy basis, where $y \in \mathbb{N}$. While the points in the Poisson case are all of the same size (corresponding to 1) we used different point sizes to indicate different natural numbers in the negative Binomial case. Alternatively, one could have used a three-dimensional picture where the third dimension reflects the size of y or one could think of a marked point process, where the mark reflects the size of y.

2.2 Definition of integer-valued trawl processes

After having defined an integer-valued, homogeneous Lévy basis, we turn to defining the trawl. Throughout the paper we will denote the Lebesgue measure by $leb(\cdot)$.

[1]Recall that a probability measure ρ is infinitely divisible, if for any $n \in \mathbb{N}$ there exists a probability measure ρ_n such that ρ can be written as the n-fold convolution of the measure ρ_n with itself, cf. Sato (1999, p. 31). In our case, ρ represents the law of L.



(a) (b)

Figure 1: (a) Point pattern associated with a realisation of a Poisson Lévy basis on $[0,1] \times \mathbb{R}$. (b) Point pattern associated with a realisation of a negative Binomial Lévy basis on $[0,1] \times \mathbb{R}$. While in the Poisson case each dot represents $y = 1$, in the negative Binomial case, the different dot sizes indicated different values of $y \in \mathbb{N}$.

Definition 5 *A* trawl *is a Borel set $A \subset [0,1] \times (-\infty, 0]$ such that $leb(A) < \infty$. Then, we set*

$$A_t = A + (0,t), \tag{4}$$

which gives the position of the trawl at time t.

Note that by working with the definition (4), we ensure that the trawl is non-anticipative. From the definition it is obvious that the size of the trawl does not change over time, i.e. we have $leb(A_t) = leb(A)$ for all t.

Typically, we choose A to be of the form

$$A = \{(x,s) : s \leq 0,\ 0 \leq x \leq d(s)\}, \tag{5}$$

where $d : (-\infty, 0] \mapsto [0,1]$ is a continuous function such that $leb(A) < \infty$. Then we define

$$d^\star = leb(A) = \int_{-\infty}^{0} \int_{0}^{d(s)} dxds = \int_{-\infty}^{0} d(s)ds. \tag{6}$$

Further,

$$A_t = A + (0,t) = \{(x,s) : s \leq t,\ 0 \leq x \leq d(s-t)\}.$$

If, in addition, d is monotonically non-decreasing, then we call A a *monotonic trawl*.

Figure 2 illustrates the construction of the trawl. At time zero, we consider the set $A = A_0$, which can be any set in $\mathbb{R}^2$, which has finite Lebesgue measure. As time progresses, we shift the trawl forward along the time axis. E.g. at time t, we consider the set $A_t = A + (0,t)$.

Example 1 *Suppose that $d(s) = \exp(\lambda s)$ for $\lambda > 0, s \leq 0$. Then $A = \{(x,s) : s \leq 0,\ 0 \leq x \leq d(s)\}$ is a monotonic trawl, and $A_t = A + (0,t) = \{(x,s) : s \leq t,\ 0 \leq x \leq \exp(\lambda(s-t))\}$.*

Now we define the new class of integer-valued trawl processes as an integer-valued Lévy basis evaluated over a trawl.

Definition 6 *A stationary integer-valued trawl (IVT) process $(Y_t)_{t \in \mathbb{R}}$ is defined as*

$$Y_t = L(A_t) = \int_{[0,1] \times \mathbb{R}} \mathbf{I}_A(x, s-t) L(dx, ds),$$

where L is defined as in Definition 3, and $A_t = A + (0,t)$ is a trawl as defined in Definition 5.

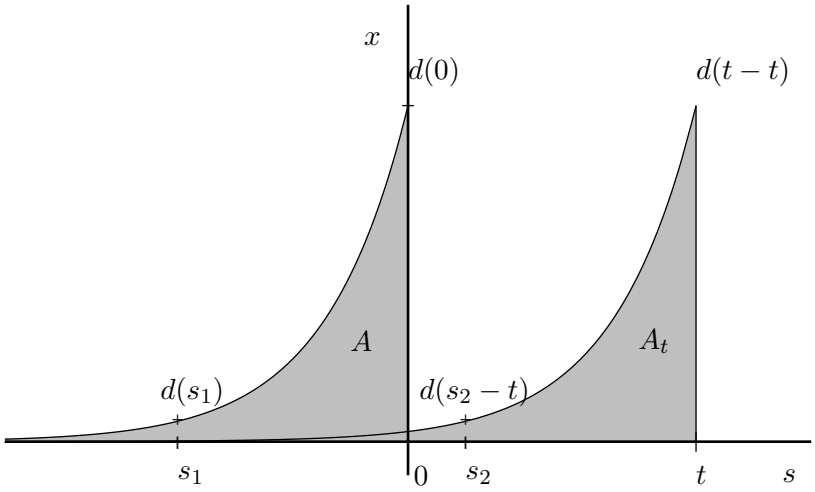


Figure 2: Illustration of a particular choice of a trawl A, which is then moved along the time axis, leading to A_t. The shape of the trawl does not change as time t progresses.

The assumption that $leb(A) < \infty$ guarantees the existence of the above integral in the sense of Rajput & Rosinski (1989). Note that by construction, an IVT process has finite variation.

Remark 2 *IVT processes take values in $\mathbb{Z}$. If one is interested in modelling count data, one needs to add the assumption that the corresponding Lévy measure ν is only supported on $\mathbb{N}$.*

Note that we can express the trawl process in terms of its associated Poisson random measure as follows

$$Y_t = \int_{\mathbb{R}\times[0,1]\times\mathbb{R}} y\mathbf{I}_A(x, s-t)N(dy, dx, ds) = \int_{\mathbb{R}\times[0,1]\times\mathbb{R}} y\mathbf{I}_{A_t}(x, s)N(dy, dx, ds).$$

Hence, intuitively, we can think of the realisation of the trawl process at time t being obtained by adding up the marks y of the points (x, s) which are contained in the set A_t (which is equivalent to saying that the points $(x, s-t)$ are in the set A). I.e. let R denote again the countable set of points stemming from the realisation of the Lévy basis. Let (y, x, s) be such a point in R. The moment time reaches s the value of the process Y changes by the amount y. As time progresses, the point (y, x, s) can leave the trawl again. In that case, Y changes by the amount y. For monotonic trawls, a point which has left the trawl will not re-enter the trawl. Note that the character of the trawl A can be learned from the autocorrelation function of Y as we will see in the following section.

Remark 3 *Note that Wolpert & Taqqu (2005) study a subclass of general trawl processes (not necessarily restricted to the (positive) integers) under the name "up-stairs" representation, which is essentially defined as "the random measure of a moving geometric figure in a higher-dimensional space", cf. Wolpert & Taqqu (2005), Section 2.3. Also, in a later paper Wolpert & Brown (2011) study so-called "random measure processes" which fall into the trawling framework.*

2.3 Properties of integer-valued trawl processes

Let us study some key properties of trawl processes.

2.3.1 Infinite divisibility and stationarity

First of all we show that an IVT process is infinitely divisible and stationary.

For any $n \in \mathbb{N}$ and $\theta_1, \ldots, \theta_n, t_1, \ldots, t_n \in \mathbb{R}$, the corresponding finite dimensional distribution is characterised through the characteristic function by

$$\begin{aligned}\mathbb{E}\left(\exp\left(i\sum_{j=1}^{n}\theta_j Y_{t_j}\right)\right) &= \mathbb{E}\left(\exp\left(i\int_{[0,1]\times\mathbb{R}}\sum_{j=1}^{n}\theta_j \mathbb{I}_A(x, s-t_j)L(dx,ds)\right)\right)\\ &= \exp\left(\int_{[0,1]\times\mathbb{R}} C\left(\sum_{j=1}^{n}\theta_j \mathbb{I}_A(x,s-t_j)\ddagger L'\right)dxds\right)\\ &= \exp\left(\int_{[0,1]\times\mathbb{R}}\int_{\mathbb{R}}\left(e^{i\sum_{j=1}^{n}\theta_j\mathbb{I}_A(x,s-t_j)y}-1\right)\nu(dy)dxds\right).\end{aligned}$$

Now we lift the measure $\nu \times leb$ from $[0,1]\times\mathbb{R}^2$ to $\mathbb{R}^n$ by defining a new measure U on $\mathbb{R}^n$ through the mapping $(x, s, y; \mathbf{t}) \mapsto (\mathbb{I}_A(x, s-t_j)y)_{j=1,\ldots,n}$, where $\mathbf{t} = (t_1, \ldots, t_n)^\top$. Let $\mathbf{z} = (z_1, \ldots, z_n)^\top$, then we have

$$\mathbb{E}\left(\exp\left(i\sum_{j=1}^{n}\theta_j Y_{t_j}\right)\right) = \exp\left(\int_{\mathbb{R}^n}(e^{i\sum_{j=1}^{n}\theta_j z_j}-1)U(d\mathbf{z})\right).$$

Note that the fact that U is a Lévy measure on $\mathbb{R}^n$ follows directly from the fact that ν is a Lévy measure on $\mathbb{R}$ and that $leb(A) < \infty$. Hence we can conclude that we have indeed an ID process.

Next, a simple change of variable argument shows that

$$\begin{aligned}&\mathbb{E}\left(\exp\left(i\sum_{j=1}^{n}\theta_j Y_{t_j}\right)\right)\\ &\qquad= \exp\left(\int_{[0,1]\times\mathbb{R}} C\left(\theta_1\mathbb{I}_A(x,u)+\sum_{j=2}^{n}\theta_j\mathbb{I}_A(x,u-(t_j-t_1))\ddagger L'\right)dxdu\right),\end{aligned}$$

i.e. the dependence on $t_1, \ldots, t_n$ is only through the differences $t_2 - t_1, \ldots, t_n - t_1$, which proves the stationarity of the processes.

2.3.2 Mixing property and ergodicity

From Definition 6 we immediately see that an IVT process is a special case of a so-called mixed moving average (MMA) process. Such processes have been introduced in the context of a stable Lévy basis by Surgailis et al. (1993), who also showed that such processes are mixing. Recall that a real-valued stochastic process $X = (X_t)_{t\in\mathbb{R}}$ on the canonical probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is *mixing* if $\mathbb{P}(A \cap S^t B) \to \mathbb{P}(A)\mathbb{P}(B)$, as $t \to \infty$; it is *weakly mixing* if $T^{-1}\int_0^T |\mathbb{P}(A \cap S^t B) - \mathbb{P}(A)\mathbb{P}(B)|dt$, as $T \to \infty$; and it is *ergodic* if $T^{-1}\int_0^T \mathbb{P}(A \cap S^t B)dt \to \mathbb{P}(A)\mathbb{P}(B)$, as $T \to \infty$, where $A, B \in \mathcal{F}$ and $S = (S^t)_{t\in\mathbb{R}}$ denotes the shift operator. Clearly a mixing process is also weakly mixing and every weakly mixing process is ergodic.

Recent work by Fuchs & Stelzer (2013) finds that (multivariate) MMA processes are mixing as long as they exist, which implies that our IVT processes are indeed mixing and hence also weakly mixing and ergodic.

Mixing properties and ergodicity are interesting properties of stochastic processes in their own right, but are particularly relevant in the context of parameter estimation. In particular, it is well-known that moment-based estimation methods are consistent if the underlying stationary stochastic process is ergodic, see e.g. Mátyás (1999).

2.3.3 Cumulants and autocorrelation structure

An immediate consequence of (2) is that the cumulant function of a trawl process is given by

$$C(\theta \ddagger Y_t) = C(\theta \ddagger L(A_t)) = leb(A)C(\theta \ddagger L'). \tag{7}$$

This is an important result, which implies that to any infinitely divisible integer-valued law π, say, there exists a stationary integer-valued trawl process having π as its one-dimensional marginal law, see Barndorff-Nielsen (2011) for more details.

We can now easily derive the cumulants of the trawl process, provided they exist. In particular, the mean and variance are given by

$$\mathbb{E}(Y_t) = leb(A)\ \mathbb{E}(L'), \qquad\qquad Var(Y_t) = leb(A)\ Var(L').$$

Further, the autocorrelation structure is given as follows. Let $h > 0$, then

$$\rho(h) = Cov(Y_t, Y_{t+h}) = leb(A \cap A_h)Var(L'). \tag{8}$$

Hence

$$r(h) = Cor(Y_t, Y_{t+h}) = \frac{leb(A \cap A_h)}{leb(A)}.$$

In the case of a monotonic trawl, we get

$$leb(A \cap A_h) = \int_{-\infty}^{0} d(s-h)ds = \int_{h}^{\infty} d(-x)dx.$$

Hence,

$$\rho(h) = \int_{h}^{\infty} d(-x)dx Var(L'),$$

and, if ρ is differentiable, then

$$\rho'(h) = -d(-h)Var(L'). \tag{9}$$

So in the case that the autocovariance function is differentiable, we can use the above equality (9) to construct a trawl process for a given differentiable autocovariance function – provided ρ' is monotone. Further, for the autocorrelation function we get

$$r(h) = Cor(Y_t, Y_{t+h}) = \frac{\int_h^\infty d(-x)dx}{\int_0^\infty d(-x)dx},$$

and hence the first derivative of the autocorrelation function is given by

$$r'(h) = \frac{-d(-h)}{\int_0^\infty d(-x)dx}.$$

2.3.4 Increments of IVT processes

Finally, let us consider the increments of an integer-valued trawl process. Note that whatever the type of trawl, we have the following representation for the increments of the process for $s < t$,

$$Y_t - Y_s = L\left(A_t \backslash A_s\right) - L\left(A_s \backslash A_t\right), \qquad \text{almost surely.} \tag{10}$$

Note that the sign "\" denotes the set difference. Due to the independence of $L\left(A_t \backslash A_s\right)$ and $-L\left(A_s \backslash A_t\right)$, we get the following result for the cumulant function of the increments

$$\begin{aligned} C(\theta \ddagger Y_t - Y_s) &= C(\theta \ddagger L\left(A_t \backslash A_s\right)) + C(-\theta \ddagger L\left(A_s \backslash A_t\right)) \\ &= leb(A_t \backslash A_s) C(\theta \ddagger L') + leb(A_s \backslash A_t) C(-\theta \ddagger L') \\ &= leb(A_{t-s} \backslash A) C(\theta \ddagger L') + leb(A \backslash A_{t-s}) C(-\theta \ddagger L'). \end{aligned} \tag{11}$$

Since $leb(A_{t-s} \backslash A) = leb(A)(1 - r(t-s)) = leb(A \backslash A_{t-s})$, we have

$$C(\theta \ddagger Y_t - Y_s) = leb(A)(1 - r(t-s)) \left[C(\theta \ddagger L') + C(-\theta \ddagger L')\right]. \tag{12}$$

We will use this result in the following to obtain the distribution of the increments of a trawl process in some concrete examples.

3 Specifying the marginal distribution

We have seen that an IVT process is fully specified by two key components: The marginal distribution associated with the Lévy seed L' and by the trawl. In the following we present some relevant choices for the stationary distribution of an IVT process, and Section 4 gives an overview on how the trawl can be specified.

3.1 The Poisson trawl process

As a good starting point for exploring the wide class of IVT processes, we choose the Poisson trawl process, which we obtain by choosing a Poisson basis for the Lévy basis L or, equivalently, by choosing the Levy seed $L' \sim \text{Poi}(v)$ for an intensity parameter $v > 0$. Then

$$Y_t = L(A_t) \sim \text{Poi}(v leb(A)), \text{ for all } t \in \mathbb{R}.$$

The intuition behind such a basic model is the following one: Suppose L is a Lévy basis on $[0,1] \times \mathbb{R}$ and $A_t \subset [0,1] \times \mathbb{R}$. If we consider a realisation of L again, then we obtain a countable set R of points (y, x, s) in $\{1\} \times [0,1] \times \mathbb{R}$, where – as before – the last two coordinates (x, s) are uniformly distributed over $[0,1] \times \mathbb{R}$ while y is the value of the basis at that point, which in the case of a Poisson basis is always equal to 1. Hence as soon as time reaches s the value of the process Y increases by 1. As time progresses, the point will eventually drop out of the trawl again, which will result in the value of the process dropping by one. This finding can be used to set up a simulation algorithm for such processes, which we do in Section 7.

Example 2 (A Poisson exponential-trawl process) *Let us consider a specific example of a monotonic trawl given by* $A_t = \{(x,s) : s \le t, 0 \le x \le e^{-\lambda(t-s)}\}$*, for* $\lambda > 0$*. The trawl thins some of the elements in the field. We call this a* Poisson exponential-trawl *model, which is motivated from the class of non-Gaussian Ornstein-Uhlenbeck (OU) processes, see e.g. Barndorff-Nielsen & Shephard (2001). It is illustrated in Figure 3. It shows the impact of the trawl* A_t *with points not in the shaded area being excluded. It highlights the times* $t = 5$ *and* $t = 7$ *to focus attention. As* Y_t *is a trawled Poisson field, at any point in time* Y_t *is marginally Poisson.*

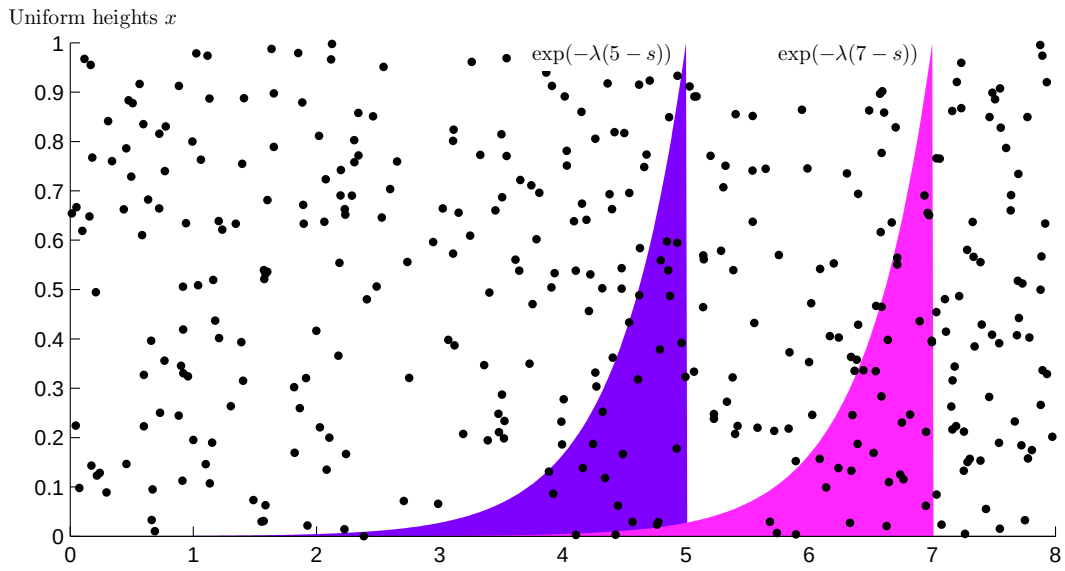


Figure 3: *Constructing a Poisson exponential-trawl process by trawling a Poisson basis with* $v = 50$ *using* $A_t = \{(x,s) : s \leq t, 0 \leq x \leq e^{-\lambda(t-s)}\}$ *for* $\lambda = 1.8$. *Snapshots are taken at* $t = 5$ *and* 7. *All points in the shaded area appear in* Y_5 *and* Y_7 *respectively.*

An immediate consequence of the results in Section 2.3 is that the increments of a Poisson trawl process follow a Skellam distribution, i.e. for $s < t$, the corresponding cumulant function, see (12), is given by

$$C(\zeta \ddagger Y_t - Y_s) = leb(A)(1 - r(t-s))v\left[\exp(i\zeta) + \exp(-i\zeta) - 2\right],$$

which is the cumulant function of the symmetric Skellam distribution, i.e.

$$Y_t - Y_s \sim \text{Skellam}(\psi_{t-s}, \psi_{t-s}), \quad \text{where} \quad \psi_{t-s} = leb(A)(1 - r(t-s))v.$$

3.2 The negative binomial trawl process

In a next step, we would like to allow for up- and downwards jumps which are not restricted to be of size one as in the Poisson/Skellam case. One possibility to overcome this restriction is to focus on the negative binomial distribution.

Throughout the paper, we denote by $NB\,(m,\theta)$ the negative binomial law with parameters $m \in \mathbb{N}$ and $\theta \in (0,1)$. Suppose that the Lévy seed satisfies $L' \sim NB(m,\theta)$, i.e. the corresponding probability mass function is given by

$$P\left\{L' = j\right\} = \frac{1}{j!}\frac{\Gamma\,(m+j)}{\Gamma\,(m)}\,(1-\theta)^m\,\theta^j, \qquad \text{for } j \in \mathbb{N}_0.$$

Remark 4 *Note that the Lévy process* $(L'_t)_{t\geq 0}$ *associated with the Lévy seed through* $L' \overset{law}{=} L'_1$ *satisfies* $L'_t \sim NB\,(mt,\theta)$. *We recall that the negative binomial Lévy process may be thought of as a compound Poisson process* $L'_t = \sum_{i=1}^{N_t} C_i$ *where* N *is a Poisson process with intensity* $v = m\,|\log\,(1-\theta)|$ *and the innovations* C_i *follow the logarithmic distribution with point probabilities*

$$\mathbb{P}(C_i = j) = \frac{1}{|\log\,(1-\theta)|}\frac{\theta^j}{j}, \qquad \textit{for } j \in \mathbb{N}. \tag{13}$$

The above representation result will prove useful in the context of simulating a negative Binomial trawl process, see Section 7. Alternatively L' *is representable through mixing the parameter* ν *of a Poisson process by endowing* ν *with the Gamma law* $\Gamma\left(m, \frac{\theta}{1-\theta}\right)$.

Recall that the cumulant function of the negative Binomial Lévy seed is given by

$$C\left(\zeta \ddagger L'\right) = m\left(\log\left(1-\theta\right) - \log\left(1-\theta e^{i\zeta}\right)\right), \tag{14}$$

which implies, in particular, that

$$Y_t = L(A_t) \sim NB(leb(A)m, \theta).$$

Also, for $s < t$, the increment $Y_t - Y_s$ is distributed as

$$Y_t - Y_s \sim \Delta NB\left(m_{t-s}, \theta, m_{t-s}, \theta\right), \quad \text{where} \quad m_{t-s} = mleb(A)(1-r(t-s)).$$

Here $\Delta NB\left(m_{t-s}, \theta, m_{t-s}, \theta\right)$ is the notation introduced in Barndorff-Nielsen, Pollard & Shephard (2012) for the difference between two independent and – in our case identically distributed – negative binomial random variables with laws $NB\left(m_{t-s}, \theta\right)$.

3.3 IVT processes with support on $\mathbb{Z}$

Many of the well-known infinitely divisible distributions which are integer-valued, such as e.g. the Poisson or negative binomial distribution have support on non-negative integers. Such distributions are suitable for modelling count data.

In some applications however, one might need distributions which have support on both the positive and negative integers. Since such a situation can be easily dealt with in our new modelling framework, we give a brief outlook on IVT processes with support on $\mathbb{Z}$ in the following.

A very natural way of tackling this problem is to work with a construction where the Lévy seed L' is modelled as the difference of two independent integer-valued Lévy subordinators $L'^{(+)}$ and $L'^{(-)}$, say, see Barndorff-Nielsen, Pollard & Shephard (2012). I.e.

$$L' = L'^{(+)} - L'^{(-)}. \tag{15}$$

Note here that due to the finiteness of the Lévy measure (which is implied by the fact that we work with integer-valued Lévy processes), we know that L' has bounded variation and, hence, has necessarily a representation of type (15). In many situations we would tend to choose the same distribution for both the upward jumps, represented by $L'^{(+)}$, and the downward jumps given by $L'^{(-)}$, although this is not a necessary requirement.

Similarly to the computations in (11) we can derive the cumulant function of L', which is given by

$$C\left(\zeta \ddagger L'\right) = C\left(\zeta \ddagger L'^{(+)}\right) + C\left(-\zeta \ddagger L'^{(-)}\right). \tag{16}$$

Let $L^{(+)}$ and $L^{(-)}$ denote the independent Lévy bases associated with the independent Lévy seeds, then we get immediately that, at least in law,

$$Y_t = L^{(+)}(A_t) - L^{(-)}(A_t).$$

3.3.1 The Skellam trawl process

Probably the most obvious example to consider is the case when $L'^{(+)}$ and $L'^{(-)}$ have Poisson(ψ^+) and Poisson(ψ^-) distribution, respectively. Then L' has the Skellam(ψ^+, ψ^-) distribution, i.e. $C(\zeta \ddagger L') = -\left(\psi^+ + \psi^-\right) + \psi^+ e^{i\zeta} + \psi^- e^{-i\zeta}$.

From equation (7) we can now deduce that the corresponding trawl process Y has a Skellam distribution, more precisely

$$Y_t \sim \text{Skellam}(leb(A)\psi^+, leb(A)\psi^-). \tag{17}$$

Example 3 *We call a trawl process Y_t a Skellam exponential-trawl process if* (17) *holds and the trawl is given by* $A_t = \{(x, s) : s \leq t, 0 \leq x \leq e^{-\lambda(t-s)}\}$.

Remark 5 *Many further classes of infinitely divisible distributions with support on the (non-negative) integers can be constructed along the lines of Barndorff-Nielsen, Pollard & Shephard (2012), Section 3.5, or by using Lévy-mixing as described in Barndorff-Nielsen, Perez-Abreu & Thorbjørnsen (2012).*

3.3.2 Markov property

It should be noted that IVT processes are not in general Markov processes. In the following, we characterise the necessary and sufficient conditions for an IVT process to be a Markov process.

Proposition 1 *Let $Y = (Y_t)_{t\in\mathbb{R}}$ denote an integer-valued trawl process as defined in Definition 6 with a trawl defined in* (5).

Then Y is a Markov process if and only if the Lévy seed associated with the Lévy basis has Poisson distribution and the trawl is determined by the exponential function, i.e. $d(x) = \exp(\lambda x)$ for $\lambda > 0, x \leq 0$.

Proof. In the case of an exponential trawl, Wolpert & Brown (2011), Theorem 3, have shown that Y is a Markov process if and only if Y has Poisson marginal law. In particular, a Poisson trawl process with exponential trawl is a birth-death process, where the birth rates are given by the intensity parameter of the Poisson distribution, and the death rates are determined by the parameter in the exponential trawl.

As soon as the trawl is not given by an exponential function (or is even non-monotonic), the resulting holding times for each state are not exponentially distributed and hence the process is not Markovian. ■

We remark that the work by Wolpert & Brown (2011) is complementary to this article since they are interested in fully characterising the class of *Markovian*, infinitely divisible, stationary, non-negative integer-valued (MISTI) processes. They mention the class of trawl processes briefly (and call them *random measure processes*), but then focus on the MISTI processes in their paper, whereas we show in this article that despite the fact that our processes are not in general Markovian, they nevertheless form a highly tractable class of stochastic processes. This makes them interesting for applications as we demonstrate in Section 8.

4 Specifying the trawl

We have discussed several examples of how we can specify an integer-valued Lévy basis, so now we will focus on specifying the trawl. We have already seen that the trawl determines the autocorrelation structure and hence is a key component in our model. As mentioned earlier, there are only two assumptions, which have to be satisfied by the trawl: First, it has to have finite Lebesgue measure. Second, since we focus on stationary processes here, we assume that the shape of the trawl does not change over time. Note that in Section 5 we will discuss how the latter assumption can be relaxed. The flexibility obtained from the trawl function sets the trawl framework apart from many traditional models for count or integer-valued data.

In order to simplify the exposition, throughout this Section, we restrict our discussions to trawls $A_t = A + (0, t)$ where $A = \{(x, s) : s \leq 0, 0 \leq x \leq d(s)\}$, for a (typically non-decreasing) function d such that $leb(A) < \infty$.

4.1 Exponential trawls and superpositions

A natural starting point is the exponential trawl, which is motivated from Ornstein-Uhlenbeck processes and was already mentioned in Example 2.

Example 4 *Suppose we have a trawl defined as the one in Figure 3, i.e.* $d(x) = \exp(\lambda x)$ *for* $\lambda > 0, x \leq 0$. *Then* $leb(A \cap A_h) = \frac{1}{\lambda}\exp(-\lambda h)$ *and, hence,* $r(h) = \exp(-\lambda h)$, *for* $h \geq 0$.

Clearly, choosing an exponential trawl leads to an autocorrelation function which can model short memory processes. A natural extension would be to consider a weighted sum of exponential functions, see the following example.

Example 5 *Let* $w_j \geq 0$, *for* $j = 1, 2, ..., J$ *and set*

$$d(t) = \sum_{j=1}^{J} w_j e^{\lambda_j t}, \quad \sum_{j=1}^{J} w_j = 1, \tag{18}$$

see e.g. Barndorff-Nielsen & Shephard (2001) for superposition type models. This allows $d(t)$ *to increase slowly for small* t *and then increase very quickly for large* t *by simply setting* $J = 2$ *and allowing* λ_1 *to be large and* λ_2 *to be small. Clearly*

$$d^* = \int_{-\infty}^{0} d(x)dx = \sum_{j=1}^{J} w_j \int_{-\infty}^{0} \exp(\lambda_j x)dx = \sum_{j=1}^{J} \frac{w_j}{\lambda_j},$$

and, hence, we have

$$\mathrm{Cor}(Y_t, Y_{t+h}) = \sum_{j=1}^{J} \frac{w_j}{\lambda_j}\exp(-\lambda_j h)\left(\sum_{j=1}^{J} \frac{w_j}{\lambda_j}\right)^{-1}.$$

A very flexible way of parametrising the trawl function, which we study in the following, is to work with a superposition of exponential trawls. Here we randomise the memory parameter λ. Such a modelling framework is motivated from the supOU processes, see e.g. Barndorff-Nielsen (2001) and Barndorff-Nielsen & Stelzer (2011). In fact, the previous example is just a special case of the general construction we study next.

Example 6 *Let*

$$d(z) = \int_0^\infty e^{\lambda z}\pi(d\lambda), \qquad \textit{for } z \leq 0,$$

for a probability measure π *on* $(0, \infty)$. *Note that if* $\pi(d\lambda)$ *is the Dirac measure for a fixed* λ, *we are back to the exponential case.*

In the case that π *is absolutely continuous with density* f_π, *we get*

$$d(z) = \int_0^\infty e^{\lambda z} f_\pi(\lambda)d\lambda.$$

Note that $d(z)$ *is increasing (non-decreasing) in* z, *hence,* $d(s-t)$ *is also increasing (non-decreasing) in* s *for* $s \leq t$. *Then the autocorrelation function is given by*

$$r(h) = Cor(Y_t, Y_{t+h}) = \frac{\int_0^\infty \frac{1}{\lambda}e^{-\lambda h}\pi(d\lambda)}{\int_0^\infty \frac{1}{\lambda}\pi(d\lambda)},$$

where we assume that $\int_0^\infty \frac{1}{\lambda}\pi(d\lambda) < \infty$.

Let us study a concrete example for the choice of π next. More precisely, we will choose a generalised inverse Gaussian (GIG) density function for f_π.

Example 7 *In the GIG case, we have*

$$f_\pi(x) = \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)} x^{\nu-1} \exp\left(-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right), \tag{19}$$

where $\nu \in \mathbb{R}$ and γ and δ are both nonnegative and not simultaneously equal to zero. Also, $K_\nu(\cdot)$ is the modified Bessel function of the third kind. Let us denote the constant in the density by $c(\gamma, \nu, \delta) := \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)}$. The trawl function is then given by

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

Also,

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

And the autocorrelation function is given by

$$r(h) = Cor(Y_t, Y_{t+h}) = \frac{K_{\nu-1}(\delta\sqrt{\gamma^2 + 2h})}{K_{\nu-1}(\delta\gamma)} \left(1 + \frac{2h}{\gamma^2}\right)^{\frac{1}{2}(1-\nu)}.$$

In the inverse Gaussian case (i.e. $\nu = \frac{1}{2}$), the trawl function has the form

$$d(z) = \left(1 - \frac{2z}{\gamma^2}\right)^{-\frac{1}{2}} \exp\left(\delta\gamma\left\{1 - \sqrt{1 - \frac{2z}{\gamma^2}}\right\}\right); \tag{20}$$

also, $leb(A) = \gamma/\delta$ and the correlation function is given by

$$r(h) = Cor(Y_t, Y_{t+h}) = \exp\left(\delta\gamma\left\{1 - \sqrt{1 - \frac{2h}{\gamma^2}}\right\}\right).$$

Throughout the paper, we will refer to a trawl function defined by (20) *as* sup-IG trawl.

In Figure 4 we illustrate the use of a Poisson Lévy basis with an exponential and a sup-IG trawl function. Figure 4 (a) shows the realisation of the point pattern coming from a Poisson Lévy basis, an exponential trawl and the corresponding sample path of the trawl process. Figure 4 (c) depicts the corresponding empirical and theoretical autocorrelation function. Using the same Poisson basis, but a different trawl function, namely a sup-IG trawl function, we show the corresponding realisation of the point pattern, the sample path in Figure 4 (b) and the autocorrelation function in Figure 4 (d). In Figure 5 we show a similar illustration, but now the Poisson Lévy basis is replaced by a negative binomial Lévy basis.

Next we turn to an example which allows for long memory.

Example 8 *Suppose π is absolutely continuous with density given by the $\Gamma(H, \alpha)$ density function, that is $\pi(\mathrm{d}\lambda) = f_\pi(\lambda)d\lambda$ with*

$$f_\pi(\mathrm{d}\lambda) = \frac{1}{\Gamma(H)} \alpha^H \lambda^{H-1} e^{-\lambda\alpha}.$$

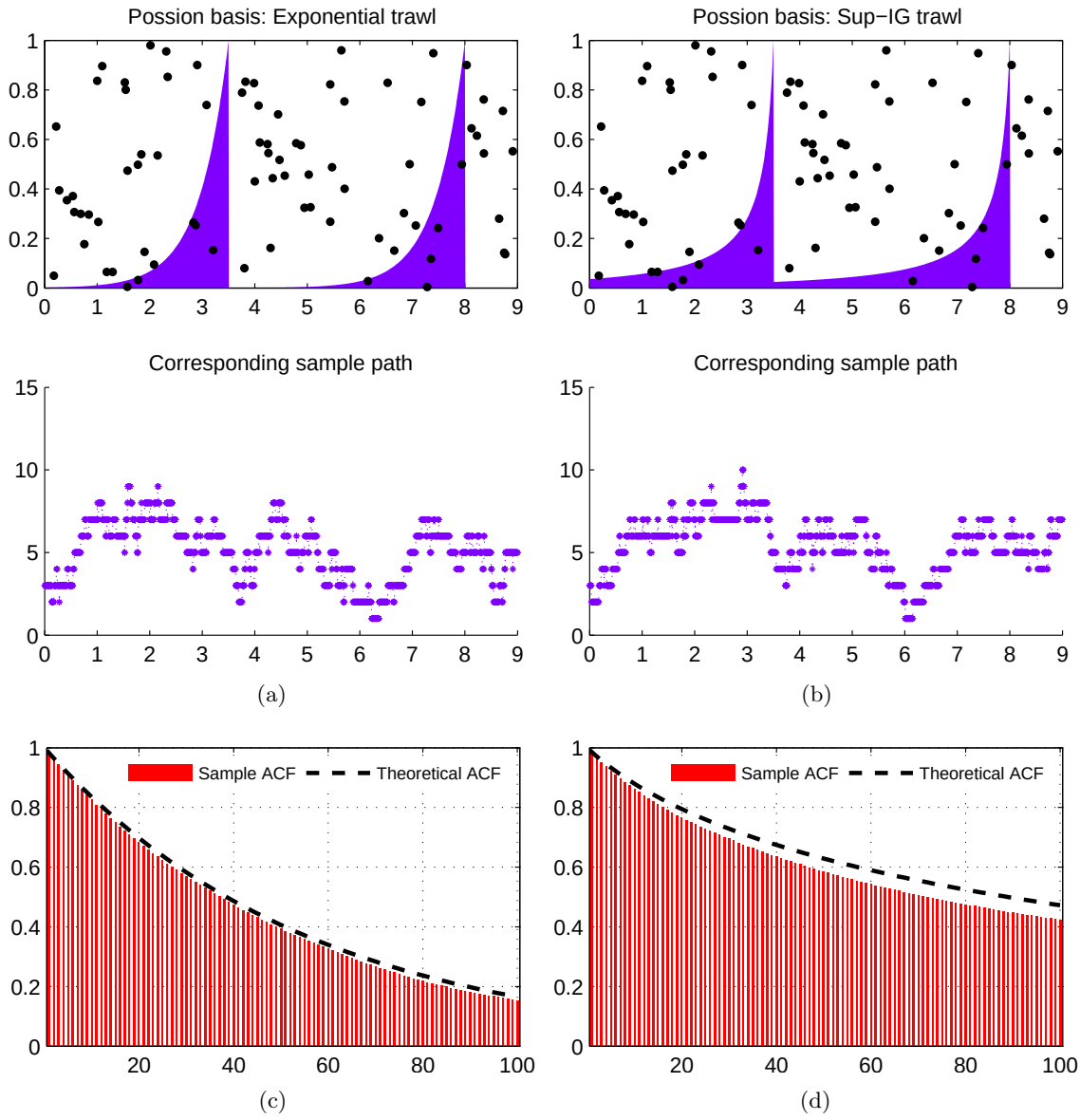


Figure 4: Poisson basis with exponential and sup-IG trawl function: (a) The upper picture shows a realisation of the Poisson Lévy basis with parameter $\upsilon = 10$ with exponential trawl with parameter $\lambda = 1.8$. The lower picture shows the sample path of the corresponding trawl process. (b) The upper picture shows a realisation of the Poisson Lévy basis with parameter $\upsilon = 10$ with sup-IG trawl, where the corresponding parameters are given by $\delta = 0.75$ and $\gamma = 0.5$. The lower picture shows the sample path of the corresponding trawl process. (c) and (d) show both the sample and the theoretical autocorrelation function for the exponential trawl and the sup-IG trawl, respectively.

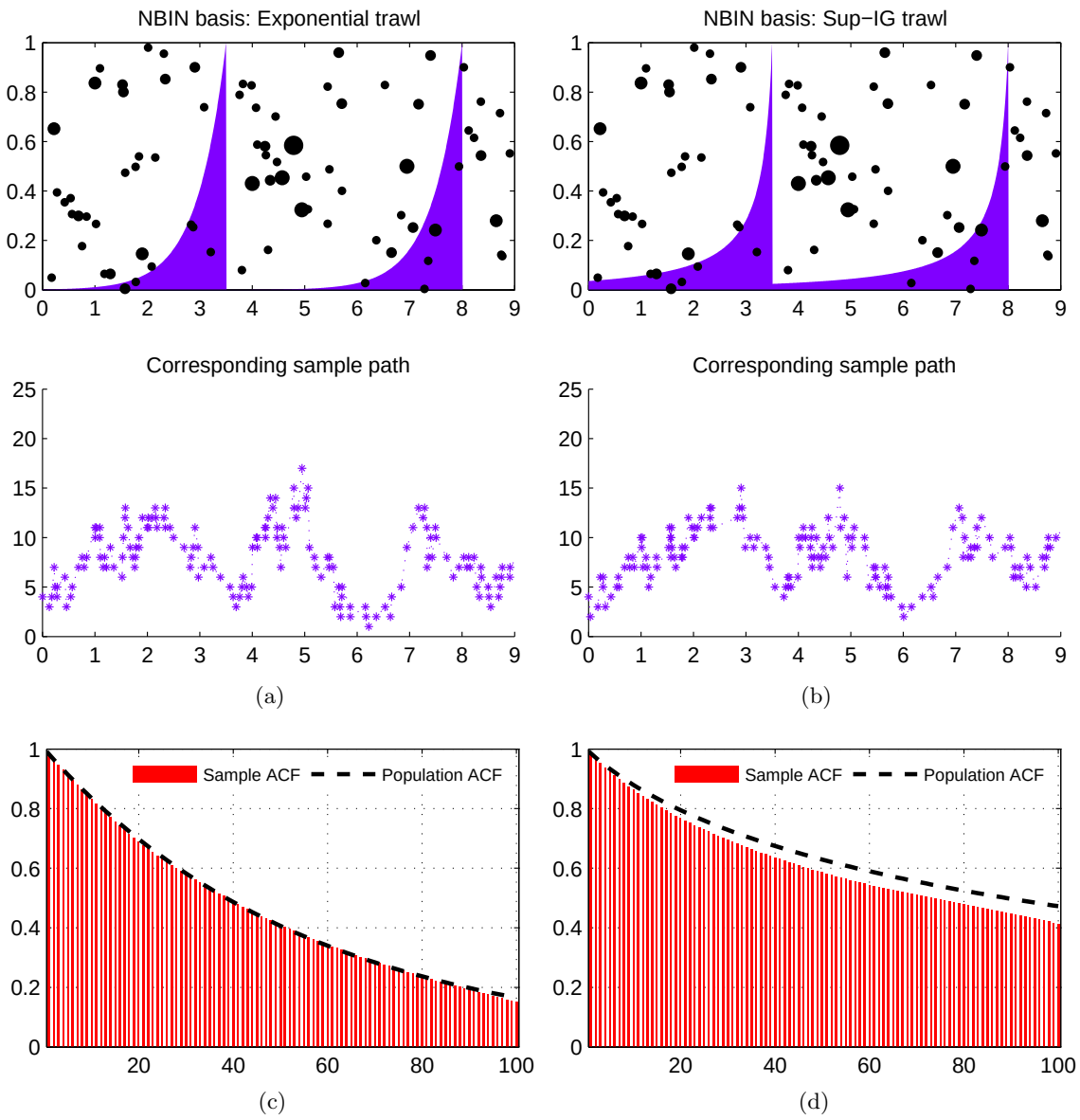


Figure 5: Negative Binomial basis with exponential and sup-IG trawl function: (a) The upper picture shows a realisation of the negative Binomial Lévy basis with parameters $m = 10$ and $\theta = 0.5$ with exponential trawl with parameter $\lambda = 1.8$. The lower picture shows the sample path of the corresponding trawl process. (b) The upper picture shows a realisation of the negative Binomial Lévy basis with parameters $m = 10$ and $\theta = 0.5$ with sup-IG trawl, where the corresponding parameters are given by $\delta = 0.75$ and $\gamma = 0.5$. The lower picture shows the sample path of the corresponding trawl process. (c) and (d) show both the sample and the theoretical autocorrelation function for the exponential trawl and the sup-IG trawl, respectively.

Recall that the Gamma density can be obtained from the GIG density, see (19)*, as the limiting case when* $\delta \to 0$*. Then*

$$d(z) = \int_0^\infty \frac{\alpha^H \lambda^{H-1}}{\Gamma(H)} e^{-\lambda\alpha\left(1-\frac{z}{\alpha}\right)} \mathrm{d}\lambda = \left(1 - \frac{z}{\alpha}\right)^{-H}, \tag{21}$$

which means that $d^* = \frac{\alpha}{H-1} < \infty$*, if* $H > 1$*. Also,* $leb(A \cap A_h) = \frac{\alpha}{H-1}\left(1 + \frac{h}{\alpha}\right)^{1-H}$*, so*

$$\mathrm{Cor}(Y_t, Y_{t+h}) = \left(1 + \frac{h}{\alpha}\right)^{1-H}. \tag{22}$$

Note that the integrated autocovariance

$$\int_0^\infty \mathrm{Cov}(Y_t, Y_{t+h})\mathrm{d}h = \int_0^\infty \lambda^{-2}\pi(\mathrm{d}\lambda) = \int_0^\infty \frac{\alpha^H \lambda^{H-3}}{\Gamma(H)} e^{-\lambda\alpha}\mathrm{d}\lambda,$$

is infinite if $H \leq 2$*, otherwise it is finite and equals*

$$\int_0^\infty \mathrm{Cov}(Y_t, Y_{t+h})\mathrm{d}h = \frac{\alpha^2 \Gamma(H-2)}{\Gamma(H)}.$$

Hence if $H \in (1, 2]$ *then this is a stationary long-memory model, while if* $H > 2$ *it is a stationary short-memory process.*

In Figure 6 we display this process in the cases of a Poisson basis and a negative binomial basis, when $\alpha = 1$*, and* $H = 1.03$*. The corresponding correlogram for this process, based upon a sample of length* $1,000,000$*, is given in the lower right of Figure 6 together with the corresponding true autocorrelation function based on (22). This shows the very slow decay in the function for longer lags. Even though the sample size of this process is massive, there is quite a large difference between the true and empirical autocorrelation function. This is not surprising as it is known that correlograms converge very slowly for long-memory processes, e.g. Hosking (1996).*[2]

4.2 A seasonal trawl function

So far we have only focused on monotonic trawl functions. However, our modelling framework also allows for seasonally fluctuating trawl functions.

Example 9 *A possible seasonal model would be*

$$d(t) = d_p(t) d_s(t),$$

where $d_p(t)$ *monotonically increases with* t *while* $d_s(t)$ *is a purely periodic seasonal effect. Notice this does not generate different average levels in different seasons. An example of this structure is where*

$$d(t) = \frac{1}{2}\exp(\lambda t)\left[\cos(at) + 1\right], \quad a = 2\pi\psi,$$

where ψ *is the period of the seasonal (hence if* $\psi = 10$ *the seasonal lasts 10 units of time). The important feature here is that* d *is not generally monotonic. This is illustrated in Figure 7 based on* $\lambda = 0.3$*. We obtain*

$$d^* = \frac{1}{2\lambda}\left(\frac{2\lambda^2 + a^2}{\lambda^2 + a^2}\right),$$

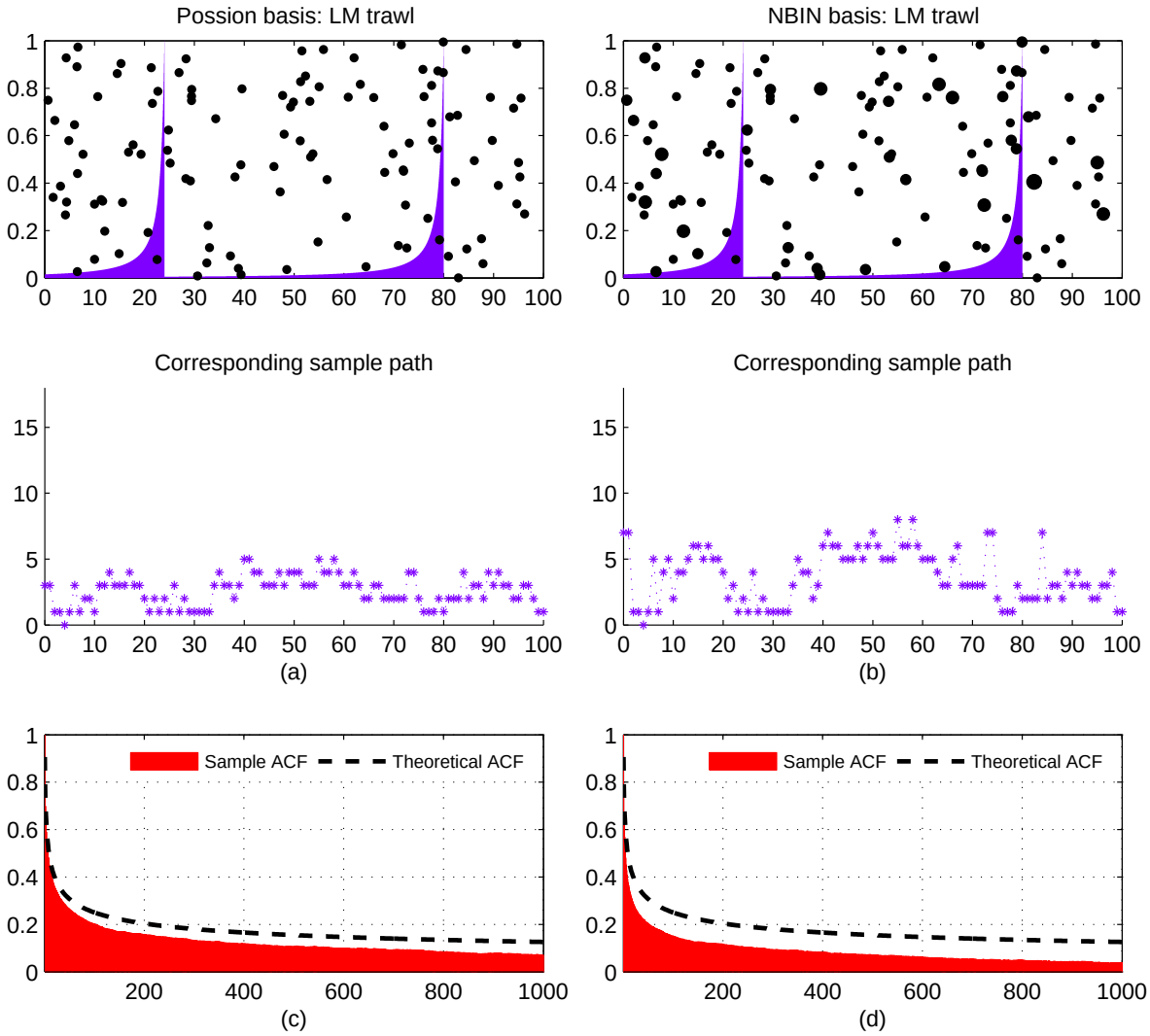


Figure 6: *Poisson and negative binomial basis with long memory trawl function: (a) The upper picture shows a realisation of the Poisson Lévy basis with parameter* $v = 1$ *with a long memory with parameter* $\alpha = 1$ *and* $H = 1.3$*. The lower picture shows the sample path of the corresponding trawl process. (b) The upper picture shows a realisation of the negative Binomial Lévy basis with parameters* $m = 1$ *and* $\theta = 0.5$*. The lower picture shows the sample path of the corresponding trawl process. (c) and (d) show both the sample and the theoretical autocorrelation function for the two cases.*

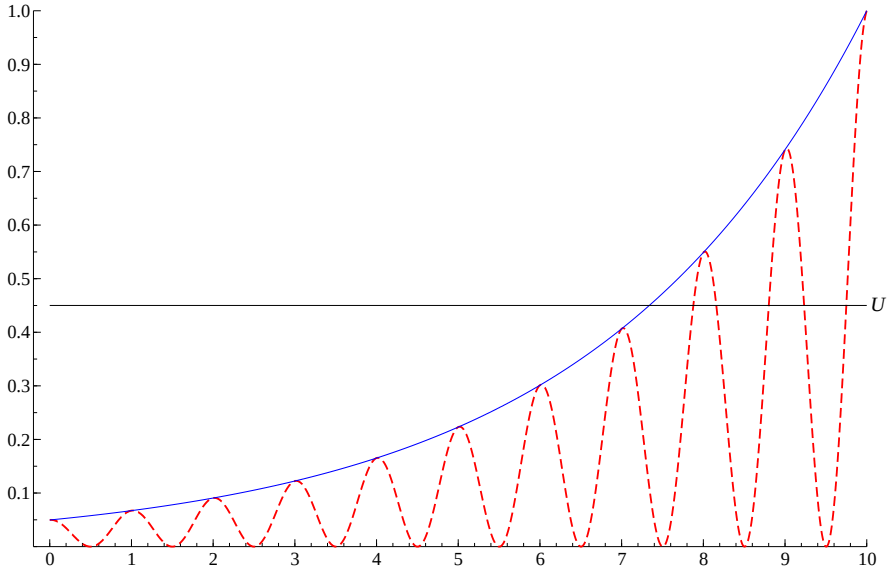


Figure 7: *Trawl function for a seasonal process using* $\lambda = 0.3$ *and the uniform height with realisation* $U = 0.45$. *In order to count in the* Y_t *process, the corresponding realisation from the Poisson field would have to fall below the dashed line. The solid line is the* $\exp(-\lambda(10-t))$ *function.*

and

$$\mathrm{Cor}(Y_t, Y_{t+h}) = \frac{e^{-\lambda h}}{2\lambda^2 + a^2}\left(\lambda^2\cos(ah) + a\lambda\sin(ah) + \lambda^2 + a^2\right).$$

Figure 8 illustrates this with $\psi = 10$ *and* $\lambda = 0.1$.

5 Volatility modulated integer-valued trawl processes

A very natural question to ask is how some form of *stochastic volatility* can be introduced in an (integer-valued) trawl process. Stochastic volatility has been studied in great detail in financial applications, see e.g. Shephard (2005), Chapter 1, and Barndorff-Nielsen & Shephard (2013) for reviews. In the classical stochastic volatility literature, one distinguishes between two types of volatility modulation: Either one introduces stochastic volatility in terms of a stochastic integrand or as a time change. While the former can be regarded as stochastically changing the amplitude of the stochastic process, the latter corresponds to modulating its intensity.

We should be aware that these two concepts are fundamentally different (unless we are in a Brownian or stable case, where the scaling property allows us to transfer one case into the other). Working with the former concept of volatility modulating is generally not suitable if one wants to stay within the class of integer-valued processes. Hence we focus on the concept of *time change*, which will ensure that we stay in the class of integer-valued processes.

The concept of stochastic time change was introduced by Bochner (1949), see also Veraart & Winkel (2010) for further references. So if $(Y_t)_{t\in\mathbb{R}}$ is an IVT processes as defined before, then we define

$$\tilde{Y}_t = Y_{T_t}, \quad t \geq 0,$$

[2]If $\mathrm{Cor}(Y_t, Y_{t+h}) \sim \lambda h^{1-H}$, then if $H \in (1,2)$ then $n^{(H-1)/2}\left\{\overline{Y} - E(Y_t)\right\}$ has a zero mean Gaussian limit. Hence $H = 1.03$ delivers a remarkably slow rate of convergence of averages to their means

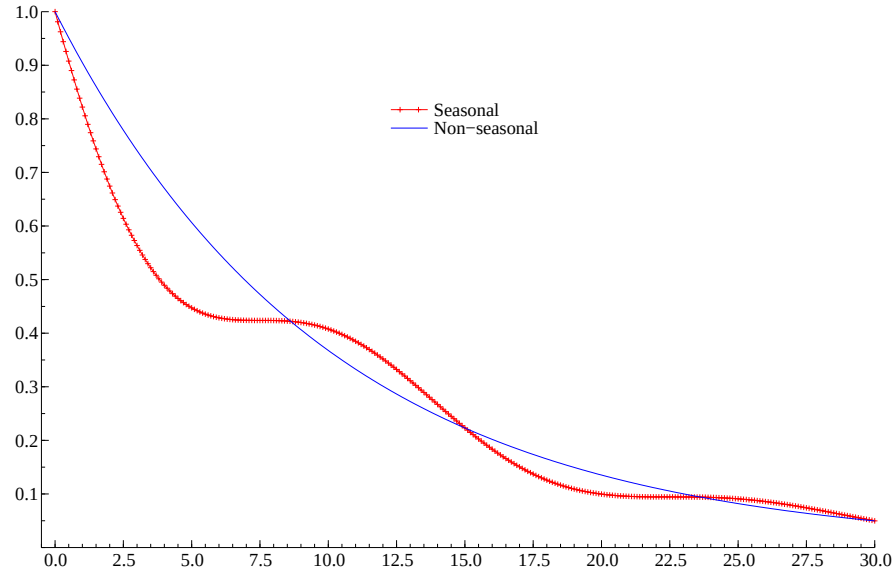


Figure 8: $\mathrm{Cor}(Y_t, Y_{t+h})$ for a seasonal model with $\psi = 10$ and $\lambda = 0.1$. The solid blue line represents the non-seasonal exponential decline, while the red line plus symbol represents the seasonal autocorrelation function.

where T is a random time change, that is an increasing stochastic process with stationary increments.

A suitable choice for the time change process is a process of the type

$$T_t = \int_0^t \tau_s ds, \quad t \geq 0,$$

where $\tau = (\tau_t)_{t\geq 0}$ denotes a stationary nonnegative process, e.g. a non-Gaussian Ornstein-Uhlenbeck process driven by a Lévy subordinator. We typically refer to τ as the *intensity process*.

Then $\tilde{Y}$ is again a stationary process and T embodies the volatility. Clearly, a time changed integer-valued process will be integer-valued itself.

Alternative ways of accounting for stochastic volatility include *extended subordination*, see Barndorff-Nielsen (2010), Barndorff-Nielsen & Pedersen (2011), *probability* and *Lévy mixing*, see Barndorff-Nielsen, Perez-Abreu & Thorbjørnsen (2012), and also *random trawls*, where the trawl is no longer a deterministic set, but random. These methods will be developed in more detail elsewhere and a recent review on this aspect and related topics can be found in Barndorff-Nielsen, Benth & Veraart (2012).

Note that stochastic volatility introduces a clustering-effect in the data. In the context of count data, one often finds an excess of zeros compared to standard distributions such as the Poisson or negative binomial distribution. Traditional models often work with zero-inflated distributions, see e.g. Johnson et al. (1992), Chapter 8, Section 2.2. However, such an effect could also be regarded as a volatility effect. E.g. when thinking of random trawls, one could easily shrink and expand the trawl such that clusters of zeros (or of any other value) can appear in the sample paths of an IVT process. This would be an instance of intensity volatility.

6 Extensions to the multivariate case

There are various ways of extending trawling to a multivariate framework, and we will describe such models in the following.

6.1 Multivariate IVT processes based on a multivariate Lévy seed

Dependence between the components of a multivariate trawl process can be introduced by dependent Lévy bases and by intersecting trawls.

One possibility of defining a multivariate trawl process $\mathbf{Y}_t = (Y_t^{(1)}, \ldots, Y_t^{(n)})'$ for $n \in \mathbb{N}$ is given as follows. In the one-dimensional definition of the Lévy measure, we could replace the Lévy basis by a vector valued Lévy basis $\mathbf{L} = (L^{(1)}, \ldots, L^{(n)})'$. I.e.

$$\mathbf{Y}_t = \left(L^{(1)}(A_t^{(1)}), \ldots, L^{(n)}(A_t^{(n)}) \right),$$

where for each component $Y^{(i)}$, we define an trawl $A^{(i)} = A_0^{(i)}$ which is assumed to have finite Lebesgue measure. Then we set $A_t^{(i)} = A^{(i)} + (0, t)$ for $i \in \{1, \ldots, n\}$. As in the univariate case we restrict our attention to Lévy bases which take values in $\mathbb{Z}^n \setminus \{\mathbf{0}\}$ or $\mathbb{N}^n$.

For applications, we typically want to allow for dependence between the individual Lévy bases. Such a dependence structure can be introduced as follows. Similarly to the one-dimensional case, we can associate an n-dimensional Lévy seed $\mathbf{L}'$ with the n-dimensional Lévy basis. Then the cumulant function of the Lévy seed is given by

$$C\{\boldsymbol{\zeta} \ddagger \mathbf{L}'\} = \int_{\mathbb{R}^n} \left(e^{i\boldsymbol{\zeta}'\mathbf{y}} - 1 \right) \nu(d\mathbf{y}), \tag{23}$$

where ν denotes the joint n-dimensional Lévy measure which is supported on the (positive) integers only. Then, for the multivariate homogeneous Lévy basis we have, as in the one-dimensional case, that

$$C\{\zeta \ddagger \mathbf{L}(d\xi, ds)\} = C\{\zeta \ddagger \mathbf{L}'\} d\xi ds.$$

Let $N(d\mathbf{y}, ds, d\xi)$ denote the n-dimensional Poisson random measure associated with $\mathbf{L}(d\xi, ds)$. Further, let $n(d\mathbf{y}, d\xi, ds) = \nu(d\mathbf{y}) d\xi ds$ denote its compensator and let $\widetilde{N} = N - n$.

Then we get the following result for the covariance between two components $1 \leq i \leq j \leq n$ for $t, h \geq 0$:

$$\rho^{i,j}(h) = Cov\left(L^{(i)}(A_t^{(i)}), L^{(j)}(A_{t+h}^{(j)}) \right) = leb\left(A^{(i)} \cap A_h^{(j)} \right) \left(\int_{\mathbb{R}} \int_{\mathbb{R}} y_i y_j \nu^{i,j}(dx_i, dx_j) \right),$$

where the joint Lévy measure for the two components is given by

$$\nu^{i,j}(d\cdot, d\cdot) = \int_{\mathbb{R}} \ldots \int_{\mathbb{R}} \nu(dy_1, \ldots, dy_{i-1}, d\cdot, dy_{i+1}, \ldots, dy_{j-1}, d\cdot, dy_{j+1}, \ldots, dy_n).$$

Also, for the auto- and cross-correlation function, we get

$$Cor\left(L^{(i)}(A_t^{(i)}), L^{(j)}(A_{t+h}^{(j)}) \right) = \frac{leb(A^{(i)} \cap A_h^{(j)}) \left(\int_{\mathbb{R}} \int_{\mathbb{R}} y_i y_j \nu^{i,j}(dy_i, dy_j) \right)}{\sqrt{leb(A^{(i)}) Var(L'^{(i)}) leb(A^{(j)}) Var(L'^{(j)})}}.$$

Remark 6 *The dependence induced by the joint Lévy measure could be modelled more explicitly by using the concept of Lévy copulas, see e.g. Cont & Tankov (2004).*

6.2 A general model

It is important to note that one can construct a more general set-up for multivariate trawl processes: Suppose $\mathbf{Y}_t = (L^{(1)}(A_t^{(1)}), \ldots, L^{(n)}(A_t^{(n)}))'$, where the trawls are given by $A_t^{(i)} = A^{(i)} + (0,t)$, for $i = 1, \ldots, n$, as above, and where $L^{(i)}$ denote univariate Lévy bases for $i = 1, \ldots, n$. Note that such a construction generally does not imply that $\mathbf{L} = (L^{(1)}, \ldots, L^{(n)})$ is a multivariate Lévy basis since the joint law might not be infinitely divisible. I.e. the dependence structure between the components might be more involved than the one induced by an n-dimensional Lévy basis.

7 Simulation study and parameter estimation

For illustration purposes, we carry out a brief simulation study, where we show how IVT processes can be simulated and how their parameters can be estimated. Throughout this section we will focus on monotonic trawls only.

7.1 Simulation algorithm for IVT processes with monotonic trawls

We have already discussed in Section 2.2 that an integer-valued Lévy basis on $[0,1]\times\mathbb{R}$ generates a random point pattern. Its realisations is a countable set R of points (y,x,s) in $\mathbb{Z}\setminus\{0\}\times[0,1]\times\mathbb{R}$. Let (N_t) denote the Poisson process with intensity $v = \int_{-\infty}^{\infty}\nu(dy)$ obtained from projecting the point pattern to the time axis. Then, associated with the arrival times $t_1, ..., t_{N_t}$ of the Poisson process N_t are the standard uniform heights $U_1, ..., U_{N_t}$.

For ease of exposition, suppose that $A = \{(x,s) : s \leq 0, 0 \leq x \leq d(s)\}$, where d is monotonically non-decreasing and $leb(A) < \infty$. Now we define $g(s) := d(-s)$, which monotonically declines as s increases. Then we will find the value R_j which solves

$$d(t_j - (R_j + t_j)) = U_j \Leftrightarrow g(R_j) = U_j,$$

that is the inverse of the function $R_j = g^{-1}(U_j)$, as this is length of time a new arrival will survive the trawl function (e.g. if U_j is high then it will not survive long).

Example 10 *If $d(s) = \exp(\lambda s)$, i.e. $g(t) = \exp(-\lambda t)$ for $\lambda > 0$, then $g^{-1}(U_j) = -\lambda^{-1}\log(U_j)$. Hence the continuous time process only jumps up at the times $\{t_j\}$ and down at the times $\{t_j + g^{-1}(U_j)\}$. This means the role of λ is to lengthen the period until the new arrival is killed.*

We call $g^{-1}(U_j)$ the "random survival period" of the jth arrival. Also, let N_t^- denote the counting process corresponding to the times $\{t_j + g^{-1}(U_j)\}$.

In the case of a Poisson trawl process, we obtain the representation

$$Y_t = N_t - N_t^-,$$

where N_t, N_t^- are not independent processes. An alternative is to write the Poisson trawl process in continuous time as

$$Y_t = Y_0 + \sum_{j=1}^{N_t} 1_{\{t_j + g^{-1}(U_j) > t\}}. \tag{24}$$

Note that in the Poisson case, it suffices to treat the behaviour of the process for $t \geq 0$. The representation (24) is central to developments of a simulation algorithm given below.

Remark 7 *The representation* (24) *shows the close connection these models have to $M/G/\infty$ queues, so-called infinite-server queues, where we can think of $g^{-1}(U_j)$ as a drawing from the "service time" distribution function G.*

For an $M/G/\infty$ queue we define the normalised integrated tail

$$G^*(x) = \frac{1}{\int_0^\infty u \mathrm{d}G(u)} \int_0^x \{1 - G(u)\} \, \mathrm{d}u,$$

then the continuous time queue length at time t, written Q_t, has the property that

$$\mathrm{Cor}(Q_t, Q_{t+h}) = 1 - G^*(h).$$

This is called Reynolds' formula. A good discussion of the statistics of this kind of queue is given by, for example, (Bingham & Pitts 1999, Section 5). Other references include Lindley (1956), Reynolds (1968) and Bartlett (1978, Chapter 6.31).

Thus the $M/G/\infty$ queue is specified in terms of the service time distribution, which implies the autocorrelation function of the queue length. In trawl processes we specify a trawl function which directly parametrises the autocorrelation function. This implies the service time distribution.

When g is not monotonic then the analogy with the $M/G/\infty$ queue breaks down.

More generally, an IVT process with monotonic trawl can be represented as

$$Y_t = Y_0 + \sum_{j=1}^{N_t} C_j 1_{\{t_j + g^{-1}(U_j) > t\}}, \tag{25}$$

where the C_j are independent, identically distributed random variables taking values in $\mathbb{Z} \setminus \{0\}$ (or in $\mathbb{N}$).

Suppose we want to simulate a sample path of an IVT process Y with representation (25) on the time interval $[0, T]$ for $T > 0$.

Algorithm 1 (Simulation algorithm for an IVT process with representation (25)**)**

1. *Generate the Poisson basis, $L(dx, ds)$, with mean measure $vdxds$ as pairs $\{t_i, U_i\}_{i=1,\ldots,tv}$. Here the t_is constitute the series of tv arrival times that are uniformly distributed on $[0, t]$, so the expected number of arrivals in each unit interval is v. Corresponding to each arrival time are the random heights, U_i, that are uniform on $[0, 1]$.*

2. *The trawl process is now constructed on a Δ-grid*

$$Y_{\Delta j} = \sum_{l=1}^{\Delta j} C_l 1_{\{g(\Delta j - l) \geq U_l\}}, \quad j = 1, \ldots, t/\Delta,$$

where $g(\Delta j - l) = \exp(-\lambda(\Delta j - l))$ for an exponential trawl (We use $\Delta = 10000$ and pretend that we observe the process on a 0.1 grid). For the sup-IG case we use $g(\Delta j - l) = d(-(\Delta j - l))$ with $d(\cdot)$ given in (20).

For the Poisson based trawl processes we set $C_l \equiv 1$. For a negative binomial trawl process, the innovations C_l follow the logarithmic distribution, see (13)*.*

7.1.1 Discretisation of an IVT process

Next let us briefly study the structure of a discretised IVT process. In particular, we will see that the discretised *Markovian* IVT processes can be related to an integer-valued AR(1) (INAR(1)) process.

Now using the same notation as in the simulation algorithm, suppose we observe the IVT process at equidistant points on a Δ-grid. Then we can express Y_t as

$$Y_t = L(A_t \cap A_{t-\Delta}) + L(A_t \setminus A_{t-\Delta}),$$

where $L(A_t \cap A_{t-\Delta})$ can be regarded as the value of the Lévy basis due to the *"survivors"* from time $t-\Delta$, and $L(A_t \setminus A_{t-\Delta})$ represents the value of the Lévy basis due to *new arrivals* in $[t-\Delta, t]$ which have not yet been killed in the interval $[t-\Delta, t]$. By definition of the Lévy basis these two terms are independent and hence the cumulant function can be written as

$$\begin{aligned} C(\theta \ddagger Y_t) &= C(\theta \ddagger L(A_t \cap A_{t-\Delta})) + C(\theta \ddagger L(A_t \setminus A_{t-\Delta})) \\ &= (leb(A_\Delta \cap A_0) + leb(A_\Delta \setminus A_0))C(\theta \ddagger L'). \end{aligned}$$

We note that the term $L(A_t \cap A_{t-\Delta})$ cannot in general be expressed in terms of $Y_{t-\Delta}$ due to the non-Markovian structure of the IVT process. However, in the case of a Markovian trawl, i.e. a Poisson trawl with $L' \sim Poi(v)$ for $v > 0$ and with exponential trawl function with parameter $\lambda > 0$, we get a more explicit result. From our previous considerations we know that the probability of "surviving" an interval of length Δ is given by $\mathbb{P}(g^{-1}(U) \leq \Delta) = g(\Delta) = \exp(-\lambda\Delta)$. Hence we can write $L(A_t \cap A_{t-\Delta}) = g(\Delta) \circ Y_{t-\Delta}$, where $\circ$ denotes the Binomial thinning operator. Now set $\Delta = 1$, then we get the following INAR(1) representation of a discretised Poisson-exponential IVT process:

$$Y_t = g(1) \circ Y_{t-1} + \epsilon_t, \quad t \in \mathbb{N},$$

where ϵ_t is independent of Y_{t-1} and the $(\epsilon_t)_{t\in\mathbb{N}}$ are i.i.d. with $C(\theta \ddagger \epsilon_t) = leb(A_1 \setminus A_0)C(\theta \ddagger L')$. More precisely, $g(1) = \exp(-\lambda)$ and $leb(A_1 \setminus A_0) = \lambda^{-1}(1 - \exp(-\lambda))$, which implies that $\epsilon_t \sim Poi(v\lambda^{-1}(1 - \exp(-\lambda)))$. It is well-known that for an INAR(1) process with thinning parameter $\alpha > 0$, the autocorrelation function is given by $\rho_j = \alpha^j$ for $j \in \mathbb{N}$. In our case this means that $\rho_j = \exp(-\lambda j)$, which is in line with our computations in Example 4.

7.2 Parameter estimation by GMM

In the following we will describe how the parameters of an IVT process can be estimated using the method of moments, see e.g. Mátyás (1999) and Hall (2005) for textbook treatments on the generalised method of moments (GMM).

Let θ_0 denote the vector of the unknown parameters from both the trawl and the Lévy basis we wish to estimate, and let Θ denote the corresponding compact parameter space satisfying $\theta_0 \in \Theta$.

Suppose we have $n \in \mathbb{N}$ equidistant observations of the trawl process Y on a Δ-grid (as described in the previous section) denoted by $(Y_\Delta, Y_{2\Delta}, \ldots, Y_{n\Delta})$. Then we define the vector

$$Z_t = (Y_{t\Delta}, Y_{(t+1)\Delta}, \ldots, Y_{(t+\tilde{n})\Delta}), \quad \text{where } t \in \{1, \ldots, n - \tilde{n}\},$$

for an $\tilde{n} \in \mathbb{N}$ with $\tilde{n} \geq 2$ representing the number of lags in the autocovariance function we will consider for the estimation. Next, we need to find a vector of measurable functions f satisfying a population moment condition of the form $\mathbb{E}[f(Z_t, \theta_0)] = 0$. More precisely, we need that θ_0 is the *unique* solution, i.e. for any $\theta \in \Theta$ we need

$$\mathbb{E}[f(Z_t, \theta)] = 0 \quad \Leftrightarrow \quad \theta = \theta_0,$$

meaning that θ_0 is *identifiable* through the moment condition. Then we estimate θ_0 as follows. We denote by $g_{n-\tilde{n}}(\theta) = (n-\tilde{n})^{-1}\sum_{t=1}^{n-\tilde{n}} f(Z_t,\theta)$ the sample moment of $\mathbb{E}[f(Z_t,\theta)]$. The GMM estimator of θ_0 is the value of θ minimising

$$g_{n-\tilde{n}}(\theta)^\top W_{n-\tilde{n}} g_{n-\tilde{n}}(\theta),$$

where $W_{n-\tilde{n}}$ denotes a positive definite weight matrix. Hall (2005) describes that a two-step iterated GMM is advantageous and we follow this procedure in the following.

Recall that an IVT process is stationary, mixing and hence ergodic. So we can deduce the following result: For fixed $t \in \mathbb{R}$ assume that $\mathbb{E}[f(Z_t,\theta)]$ is a function only depending on θ. If θ_0 is identifiable through the moment condition, then the corresponding GMM estimator is consistent.

Now, let us consider some examples and derive the relevant identifying moment conditions. We will identify the parameters of the trawl through a suitable number of lags in the autocorrelation function and we will estimate the parameters from the Lévy basis through suitable sample moments of the trawl process.

First, we consider the exponential trawl, which is characterised by one parameter, denoted by $\lambda > 0$. Referring back to Example 4, it is clear that we can estimate λ by

$$\hat{\lambda} = \log(\hat{r}_h)\frac{\Delta}{0.1 \cdot h}.$$

In the case of the sup-IG trawl as in Example 7, which depends on two nonnegative parameters denoted by γ, δ, we can utilize that the autocorrelation function is given by

$$r(h) = r(h;\delta,\gamma) = \exp\left(\delta\gamma\left\{1 - \sqrt{1 - \frac{2h}{\gamma^2}}\right\}\right),$$

and find the γ and δ that minimize

$$\sum_{h=1}^{H}\left[r(\tfrac{\Delta}{0.1\cdot h};\delta,\gamma) - \hat{r}_h\right]^2, \quad \text{for } H \geq 2.$$

Next, we study how we can identify the parameters of two different Lévy basis: a Poisson basis and a negative binomial basis.

We start with a Poisson Lévy basis with $L' \sim Poi(v)$ for $v > 0$. Recall that $\mathbb{E}(Y_t) = leb(A)\mathbb{E}(L')$. If $leb(A)$ is identifiable from the autocovariance function, then we can identify the mean of the Lévy basis from the moment condition $v = \mathbb{E}(L') = \mathbb{E}(Y_t)leb(A)^{-1}$.

We first consider the estimation of a Poisson trawl process with exponential trawl. In Table 1 we present the simulation results based on 1000 repetitions with $v = 50$ and $\lambda = 1.8$ for various sample sizes. Here v is estimated using the fact that $\mathbb{E}(Y_t) = leb(A)\mathbb{E}(L') = v/\lambda$ for this model. So $\hat{v}$ is simply the sample average of the process multiplied by $\hat{\lambda}$. It is interesting to note that quite a long sample is needed to estimate λ precisely.

Table 1: Estimation results for a Poisson trawl process with exponential trawl

		$\hat{v}$			$\hat{\lambda}$	
Nobs	ave	bias	rmse	ave	bias	rmse
100	50.0126	0.0126	3.0436	2.2861	0.4861	0.9626
250	50.0205	0.0205	1.8733	1.9734	0.1734	0.4960
500	50.0722	0.0722	1.3850	1.8828	0.0828	0.3237
750	50.0465	0.0465	1.1321	1.8557	0.0557	0.2603
1000	50.0325	0.0325	0.9654	1.8417	0.0417	0.2207
2000	50.0359	0.0359	0.6234	1.8243	0.0243	0.1532
4000	50.0122	0.0122	0.3635	1.8124	0.0124	0.1082

Now for the same Levy basis we replace the exponential trawl with the sup-IG version. Then we have $\mathbb{E}(Y_t) = leb(A)\mathbb{E}(L') = \gamma\delta^{-1}v$. So $\hat{v}$ is estimated using the sample average of the trawl process multiplied by $\hat{\delta}\hat{\gamma}^{-1}$. In Table 2 we present the simulation results based on 1000 repetitions with $v = 50$, $\gamma = 2$, $\delta = 2$ and $H = 5$ for various sample sizes.

Table 2: Estimation results for a Poisson trawl process with a sup-IG trawl

		$\hat{v}$			$\hat{\gamma}$			$\hat{\delta}$	
Nobs	ave	bias	rmse	ave	bias	rmse	ave	bias	rmse
100	49.987	-0.013	3.159	2.405	0.405	3.093	4.704	2.704	10.36
250	50.012	0.012	2.054	1.868	-0.132	0.645	2.256	0.256	1.640
500	49.961	-0.039	1.477	1.913	-0.087	0.237	2.087	0.087	0.244
750	49.992	-0.008	1.197	1.948	-0.052	0.186	2.052	0.052	0.192
1000	50.000	0.000	1.060	1.959	-0.041	0.160	2.042	0.042	0.166

For the two simulation scenarios studied above the processes could only increase or decrease by one. Next we consider a negative binomial Lévy basis, characterised by two parameters $m \in \mathbb{N}$ and $\theta \in (0, 1)$, which results in a more flexible model. Now we need two moment condition to identify m and θ in addition to identifying the parameters of the trawl function. Here we can estimate m and θ from the first two moments of the process. Since $\mathbb{E}(Y_t) = leb(A)\mathbb{E}(L')$ and $Var(Y_t) = leb(A)\,Var(L')$ with $\mathbb{E}(L') = \theta m(1-\theta)^{-1}$ and $Var(L') = pm(1-\theta)^{-2}$, we just need to match the scaled sample mean and variance of the process to the mean and variance of the negative Binomial distribution. In Table 3 we present the simulation results for the negative binomial exponential-trawl process based on 1000 repetitions with $v = 50$, $m = \nu/|\log(1-\theta)| \simeq 41.52$, $\theta = 0.7$, $\lambda = 1.8$ and $H = 5$ for various sample sizes.

Table 3: Estimation results for a negative binomial trawl process with an exponential trawl

		$\hat{m}$			$\hat{\theta}$			$\hat{\lambda}$	
Nobs	ave	bias	rmse	ave	bias	rmse	ave	bias	rmse
300	60.3	18.8	45.0	0.681	-0.0192	0.101	2.326	0.526	0.834
400	59.0	17.5	41.0	0.686	-0.0143	0.074	2.299	0.499	0.876
500	56.1	14.5	35.5	0.687	-0.0127	0.070	2.243	0.443	0.792
1000	48.9	7.41	23.2	0.695	-0.0051	0.038	2.021	0.221	0.636
2000	44.8	3.26	12.7	0.697	-0.0030	0.024	1.888	0.088	0.376
4000	43.3	1.80	8.22	0.698	-0.0020	0.017	1.847	0.047	0.249
8000	42.4	0.83	5.34	0.699	-0.0006	0.012	1.824	0.024	0.165
16000	41.8	0.25	3.67	0.700	-0.0000	0.009	1.808	0.008	0.113

Finally, in Table 4 we present the simulation results for a negative binomial sup-IG process based on 1000 repetitions with $v = 50$, $m = \nu/|\log(1-\theta)| \simeq 41.52$, $\theta = 0.7$, $\gamma = 2$, $\delta = 4$ and $H = 5$ for various sample sizes.

Table 4: Estimation results for a negative binomial trawl process with a sup-IG trawl

	$\hat{m}$			$\hat{\theta}$			$\hat{\gamma}$		$\hat{\delta}$	
Nobs	ave	bias	rmse	ave	bias	rmse	ave	rmse	ave	rmse
300	54.5	13.01	31.6	0.679	-0.0215	0.064	1.929	0.386	4.059	0.608
400	51.4	9.84	25.0	0.683	-0.0170	0.054	1.933	0.324	4.032	0.169
500	49.7	8.13	21.4	0.685	-0.0147	0.048	1.942	0.289	4.028	0.150
1000	44.8	3.25	13.1	0.694	-0.0057	0.034	1.982	0.206	4.008	0.107
2000	42.8	1.32	8.47	0.698	-0.0024	0.023	1.998	0.147	4.000	0.077
4000	42.2	0.71	5.86	0.699	-0.0012	0.016	1.998	0.106	4.001	0.055
8000	41.8	0.27	4.02	0.700	-0.0003	0.011	2.000	0.074	4.000	0.039
16000	41.6	0.06	2.83	0.700	0.0000	0.008	2.001	0.052	3.999	0.027

Overall, it appears that our moment-based estimation method works quite well. In future research it will be interesting to explore the possibility of likelihood-based estimation methods as well.

8 Empirical Illustration

As an empirical illustration we apply our trawl process to high frequency stock market data. Since most stocks are traded in prices living on a lattice, our integer-valued processes constitute a promising building block for such data. However, since asset prices themselves rarely behave like stationary processes, we rather focus on the corresponding high frequency *spread* data in our empirical study. Recall that the spread is the distance from the bid price to the ask price for an asset. Specifically, we consider two American stocks from the S&P 600 smallcap index. The stocks are Lindsay Corporation (ticker: LNN) and Navigant Consulting, Inc. (ticker: NCI) and they were selected randomly from the index. The high frequency bid and ask data used for constructing the time series of spreads have been pre-processed using the methods described in Barndorff-Nielsen et al. (2009). We study the high frequency data for both stocks for one day each using observations between 10am and 4pm, which ignores the effects right after market opening. For LNN we have 2499 observations and for NCI we have 909 observations in total. Note that we count observations as follows: An observation is generated when a new set of quote arrives, but it is only counted as a new observation if it alters the spread. So, in between observations, the spread stays constant as indicated in our plots which demonstrate that the sample path of the spread mimics a step function. As such we can think of the spread process as a continuous-time process which takes values in the natural numbers.

In Figures 9 and 10 we present the results of fitting the negative binomial trawl process with sup-IG trawl to this data. We picked a day for each ticker where the process seems to fit the data quite well. We estimated the model parameter using the moment matching method that we also used in the simulation section. Note that the spread data were not observed at equidistant time intervals, but we have new observations as soon as a quote arrives which alters the spread. While this is not a concern for estimating the mean and the variance of the trawl process, for this particular implementation of the estimation method we do need equidistant data for using the information coming from the (sample) autocovariance function. Hence we sampled each data set at equidistant time intervals (30 second intervals) using the previous tick method, where we used the most recent spread which was recorded, cf. Barndorff-Nielsen et al. (2009) for details.

Let us look at the results for LNN first. We study the data on 4 October 2012, where we have 2499 observations in total between 10am and 4pm. In Figure 9 we provide three plots. First, the evolution of the spread (in cents) is shown. Underneath the plot we report the parameter

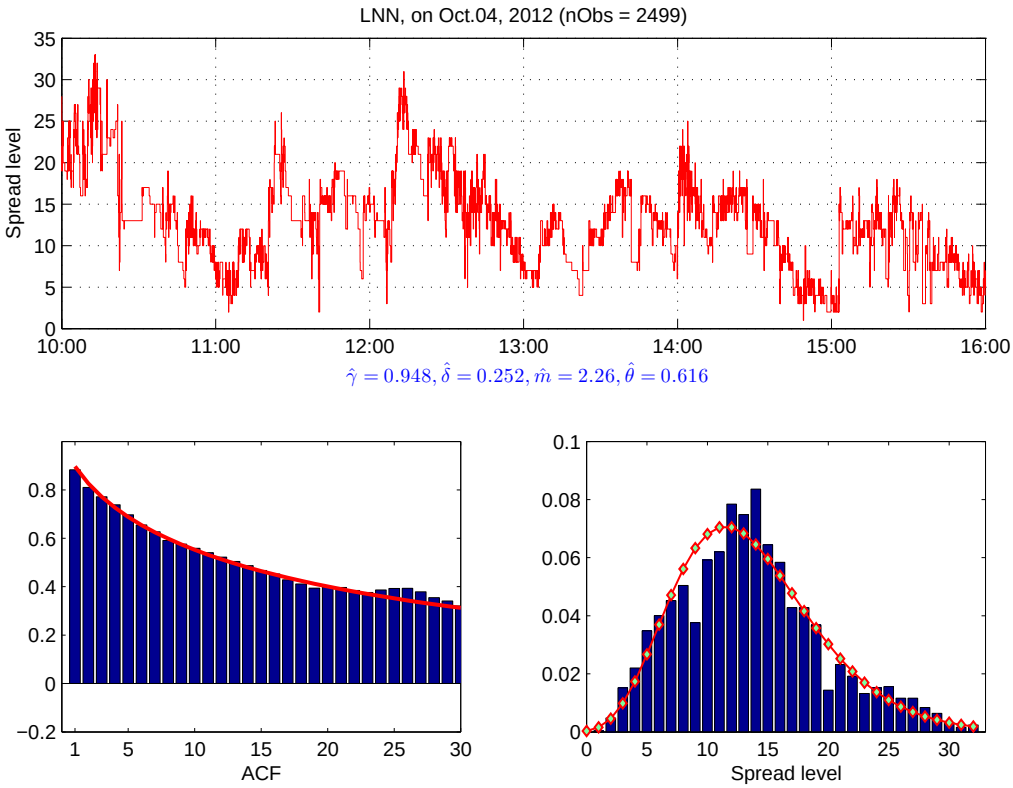


Figure 9: *Estimation of a negative binomial sup-IG trawl process for Lindsay Corporation (ticker: LNN). The first graph shows the spread level of LNN on 4 October 2012 between 10am and 4pm. Overall we have 2499 observations. The empirical and estimated autocorrelation function and the corresponding probability mass functions are compared in the second and third picture, respectively.*

estimates from our negative binomial trawl process with sup-IG trawl. The two smaller plots compare the sample ACF with the estimated ACF (where one lag corresponds to an interval of length 30 seconds) and sample histogram and with the estimated probability mass function (pmf).

We repeat the exercise for NCI, where we have 909 observations from 12 October 2012. We report the corresponding results in Figure 10.

It is interesting to note that while both processes live on a grid, LNN ranges over more than five times the number of distinct values than NCI does. In both cases the model fit appears to be quite good, especially for the LNN data. Hence we conclude that IVT processes are not only interesting from a mathematical point of view, but also represent an interesting alternative to traditional (discrete-time) time series models for count data. Due to the fact that IVT processes model the serial correlation and the marginal distribution independently of each other, they constitute a flexible class of stochastic processes. In our application to high frequency spread data, we found that an IVT process specified through a negative Binomial Lévy basis together with a sup-IG trawl function can describe our data well.

9 Conclusion

This paper has introduced the class of integer-valued trawl processes for modelling serially dependent and integer-valued data in a continuous-time framework. In terms of model building there are two key components: First, the Lévy basis provides the source of randomness deter-

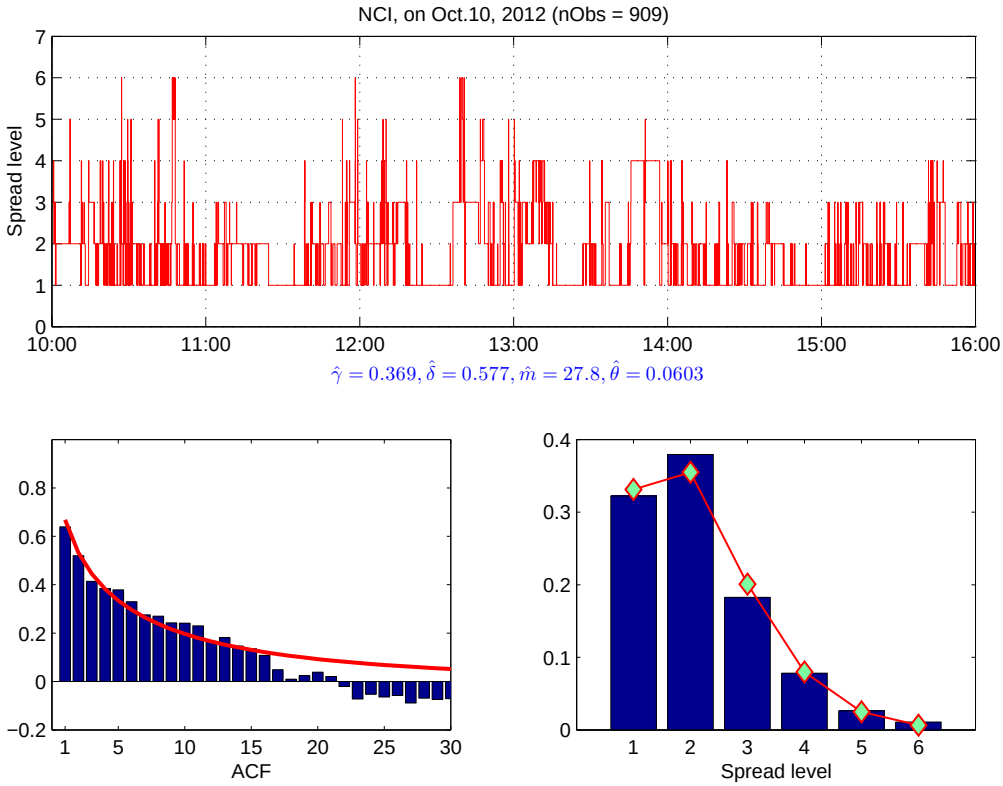


Figure 10: *Estimation of a negative binomial sup-IG trawl process for Navigant Consulting, Inc. (ticker: NCI). The first graph shows the spread level of NCI on 10 October 2012 between 10am and 4pm. Overall we have 909 observations. The empirical and estimated autocorrelation function and the corresponding probability mass functions are compared in the second and third picture, respectively.*

mining the marginal distribution of the model. All integer-valued infinitely distributions fall into our modelling framework starting from the Poisson and the negative binomial distribution to more advanced classes obtained as the marginal law from time-changing a Poisson process with a subordinator or by Lévy-mixing of an integer-valued infinite divisible distribution. Second, the choice of the deterministic function specifying the trawl is of key importance and sets the modelling framework apart from many traditional models for count data. We have seen that the trawl determines the autocorrelation function of the process and, vice-versa, the autocorrelation function determines the integrated trawl function. Here we can allow for very flexible classes of trawl functions e.g. exponential functions and superpositions thereof, which are motivated from Ornstein-Uhlenbeck and supOU processes. In addition, one could also allow for seasonal behaviour in the trawl function, which leads to non-monotonic trawls and autocorrelation functions. That said, note that the autocorrelation structure and the marginal distribution will be modelled independently of each other, which is a considerable advantage of the trawling framework compared to traditional models for count data, and we can associate a trawl process with any infinitely divisible (integer-valued) distribution.

The basic trawl framework can be extended to allow for volatility clusters. In order to ensure that one stays within the class of integer-valued processes, we have discussed that the concept of time-change is suitable. Alternatively, the concept of Lévy-mixing or random trawls can also be used for volatility modulation of an IVT process.

Multivariate trawl processes can be constructed in a very straightforward fashion, where we can allow for dependence coming both from the corresponding Lévy basis and from intersecting

trawls.

In our simulation study, we have seen that we can estimate the parameters of a given IVT process using a generalised method of moments. We have applied our estimation method in the context of high frequency financial data and obtained promising empirical results.

A Detailed computations for some of the examples

A.1 Details for Example 4

In the following we provide the computations for the exponential case in more detail. Let $d(x) = \exp(\lambda x)$, for $\lambda > 0$. Note that $leb(A) = leb(A_0) = \int_{-\infty}^{0} d(s)ds = \int_{-\infty}^{0} \exp(\lambda s)ds = \frac{1}{\lambda} < \infty$. Then $d(s-t) = \exp(\lambda(s-t)) = \exp(-\lambda(t-s))$ and the autocorrelation function is given by $r(h) = Cor(Y_t, Y_{t+h}) = \exp(-\lambda h)$.

A.2 Details for Examples 6 and 7

Now we present the detailed computations for the case where we randomise the memory parameter λ and work with a superposition type model. I.e. $d(z) = \int_0^\infty e^{\lambda z}\pi(d\lambda)$, for a probability measure π on $(0,\infty)$. In the case that π is absolutely continuous with density f_π, we get $d(z) = \int_0^\infty e^{\lambda z} f_\pi(\lambda)d\lambda$. Then

$$\begin{aligned} leb(A) = leb(A_0) &= \int_{-\infty}^{0} d(z)dz = \int_{-\infty}^{0}\int_0^\infty e^{\lambda z}\pi(d\lambda)dz = \int_0^\infty \int_{-\infty}^{0} e^{\lambda z}dz\pi(d\lambda) \\ &= \int_0^\infty \frac{1}{\lambda}\pi(d\lambda). \end{aligned}$$

Further, from the monotonicity of d, we get

$$\begin{aligned} Cov(Y_t, Y_{t+h}) = leb(A_0 \cap A_h) &= \int_{-\infty}^{0} d(s-h)ds = \int_{-\infty}^{0}\int_0^\infty e^{-\lambda h}e^{\lambda s}\pi(d\lambda)ds \\ &= \int_0^\infty e^{-\lambda h}\int_{-\infty}^{0} e^{\lambda s}ds\pi(d\lambda) = \int_0^\infty \frac{1}{\lambda}e^{-\lambda h}\pi(d\lambda). \end{aligned}$$

Then

$$r(h) = Cor(Y_t, Y_{t+h}) = \frac{\int_0^\infty \frac{1}{\lambda}e^{-\lambda h}\pi(d\lambda)}{\int_0^\infty \frac{1}{\lambda}\pi(d\lambda)}.$$

Let us focus on the GIG case, where we have

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

and $K_\nu(\cdot)$ is the modified Bessel function of the third kind. Let us denote the constant in the density as $c(\gamma,\nu,\delta) := \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)}$. The trawl function is then given by

$$\begin{aligned} d(z) &= \int_0^\infty e^{xz}\pi(dx) = \int_0^\infty e^{xz}f_\pi(x)dx \\ &= \int_0^\infty e^{xz}\frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)}x^{\nu-1}\exp\left(-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right)dx \\ &= \int_0^\infty \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)}x^{\nu-1}\exp\left(-\frac{1}{2}(\delta^2 x^{-1} + (\gamma^2 - 2z)x)\right)dx \end{aligned}$$

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

Then

$$leb(A) = \int_0^\infty \frac{1}{x}\pi(dx) = \int_0^\infty \frac{1}{x} f_\pi(x)dx = \int_0^\infty \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)} x^{(\nu-1)-1}\exp\left(-\frac{1}{2}(\delta^2 x^{-1}+\gamma^2 x)\right)dx$$
$$= \frac{c(\gamma,\nu,\delta)}{c(\gamma,\nu-1,\delta)} = \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)}\frac{2K_{\nu-1}(\delta\gamma)}{(\gamma/\delta)^{\nu-1}} = \frac{(\gamma/\delta)K_{\nu-1}(\delta\gamma)}{K_\nu(\delta\gamma)}.$$

Also,

$$\int_0^\infty \frac{1}{x}e^{-hx}\pi(dx) = \int_0^\infty \frac{1}{x}e^{-hx} f_\pi(x)dx$$
$$= \int_0^\infty \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)} x^{(\nu-1)-1}\exp\left(-\frac{1}{2}(\delta^2 x^{-1}+(\gamma^2+2h)x)\right)dx$$
$$= \frac{c(\gamma,\nu,\delta)}{c(\sqrt{\gamma^2+2h},\nu-1,\delta)} = \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)}\frac{2K_{\nu-1}(\delta\sqrt{\gamma^2+2h})}{(\sqrt{\gamma^2+2h}/\delta)^{\nu-1}}$$
$$= \frac{K_{\nu-1}(\delta\sqrt{\gamma^2+2h})}{K_\nu(\delta\gamma)}\delta^{-1}\frac{\gamma^\nu}{\sqrt{\gamma^2+2h}^{\nu-1}}$$
$$= \frac{K_{\nu-1}(\delta\sqrt{\gamma^2+2h})}{K_\nu(\delta\gamma)}\frac{\gamma}{\delta}\left(1+\frac{2h}{\gamma^2}\right)^{\frac{1}{2}(1-\nu)}.$$

Hence

$$r(h) = Cor(Y_t, Y_{t+h}) = \frac{K_{\nu-1}(\delta\sqrt{\gamma^2+2h})}{K_{\nu-1}(\delta\gamma)}\frac{\gamma^{\nu-1}}{\sqrt{\gamma^2+2h}^{\nu-1}}$$
$$= \frac{K_{\nu-1}(\delta\sqrt{\gamma^2+2h})}{K_{\nu-1}(\delta\gamma)}\left(1+\frac{2h}{\gamma^2}\right)^{\frac{1}{2}(1-\nu)}.$$

The corresponding results for the inverse Gaussian case, i.e. when $\nu = \frac{1}{2}$, follow immediately from the above results by noting that $K_{\frac{1}{2}}(x) = K_{-\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2}}x^{-\frac{1}{2}}e^{-x}$.

Acknowledgement

We wish to thank the associate editor and two anonymous referees for their valuable comments. A. E. D. Veraart acknowledges financial support by a Marie Curie FP7 Integration Grant within the 7th European Union Framework Programme and by CREATES (DNRF78), funded by the Danish National Research Foundation.

References

Barndorff-Nielsen, O. E. (2001), 'Superposition of Ornstein–Uhlenbeck type processes', *Theory of Probability and its Applications* **45**, 175–194.

Barndorff-Nielsen, O. E. (2010), Lévy bases and extended subordination. Thiele Research Report 2010–12.

Barndorff-Nielsen, O. E. (2011), 'Stationary infinitely divisible processes', REBRAPE *Brazilian Journal of Probability and Statistics* **25**(3), 294–322. Invited paper.

Barndorff-Nielsen, O. E., Benth, F. E. & Veraart, A. E. D. (2011), Ambit processes and stochastic partial differential equations, *in* G. Di Nunno & B. Øksendal, eds, 'Advanced Mathematical Methods for Finance', Springer, Berlin, pp. 35–74.

Barndorff-Nielsen, O. E., Benth, F. E. & Veraart, A. E. D. (2012), Recent advances in ambit stochastics. Preprint available on arXiv: arXiv:1210.1354.

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

Barndorff-Nielsen, O. E. & Pedersen, J. (2011), 'Meta–times and extended subordination', *Theory of Probability and its Applications* **56**, 398–407.

Barndorff-Nielsen, O. E., Perez-Abreu, V. & Thorbjørnsen, S. (2012), Lévy mixing. Preprint.

Barndorff-Nielsen, O. E., Pollard, D. G. & Shephard, N. (2012), 'Integer-valued Lévy processes and low latency financial econometrics', *Quantitative Finance* **12**(4), 587–605.

Barndorff-Nielsen, O. E. & Schmiegel, J. (2007), Ambit processes: with applications to turbulence and cancer growth, *in* F. Benth, G. Di Nunno, T. Lindstrøm, B. Øksendal & T. Zhang, eds, 'Stochastic Analysis and Applications: The Abel Symposium 2005', Springer, Heidelberg, pp. 93–124.

Barndorff-Nielsen, O. E. & Shephard, N. (2001), 'Non–Gaussian Ornstein–Uhlenbeck–based models and some of their uses in financial economics (with discussion)', *Journal of the Royal Statistical Society Series B* **63**, 167–241.

Barndorff-Nielsen, O. E. & Shephard, N. (2013), *Lévy–driven volatility models*, Cambridge University Press. To appear.

Barndorff-Nielsen, O. E. & Stelzer, R. (2011), 'Multivariate SupOU processes', *Annals of Applied Probability* **21**(1), 140–182.

Bartlett, M. S. (1978), *An Introduction to Stochastic Processes, with Special Reference to Methods and Applications*, 3 edn, Cambridge University Press, Cambridge.

Bingham, N. H. & Pitts, S. M. (1999), 'Non-parametric estimation for the M/G/∞ queue', *Annals of the Institute of Statistical Mathematics* **51**, 71–97.

Bochner, S. (1949), 'Diffusion equation and stochastic processes', *Proceedings of the National Academy of Sciences of the United States of America* **35**, 368–370.

Cameron, A. C. & Trivedi, P. K. (1998), *Regression analysis of count data*, Vol. 30 of *Econometric Society Monographs*, Cambridge University Press, Cambridge.

Cont, R. & Tankov, P. (2004), *Financial Modelling With Jump Processes*, Financial Mathematics Series, Chapman & Hall, Boca Raton, Florida, USA.

Cui, Y. & Lund, R. (2009), 'A new look at time series of counts', *Biometrika* **96**(4), 781–792.

Davis, R. A., Wang, Y. & Dunsmuir, W. T. M. (1999), Modeling time series of count data, *in* 'Asymptotics, nonparametrics, and time series', Vol. 158 of *Statistics: Textbooks and Monographs*, Dekker, New York, pp. 63–113.

Davis, R. A. & Wu, R. (2009), 'A negative binomial model for time series of counts', *Biometrika* **96**(3), 735–749.

Fuchs, F. & Stelzer, R. (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.

Hall, A. R. (2005), *Generalized method of moments*, Advanced Texts in Econometrics, Oxford University Press, Oxford.

Hosking, J. R. M. (1996), 'Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series', *Journal of Econometrics* **73**, 261–286.

Jacobs, P. A. & Lewis, P. A. W. (1978*a*), 'Discrete time series generated by mixtures. I. Correlational and runs properties', *Journal of the Royal Statistical Society. Series B. Methodological* **40**(1), 94–105.

Jacobs, P. A. & Lewis, P. A. W. (1978*b*), 'Discrete time series generated by mixtures. II. Asymptotic properties', *Journal of the Royal Statistical Society. Series B. Methodological* **40**(2), 222–228.

Johnson, N. L., Kotz, S. & Kemp, A. W. (1992), *Univariate discrete distributions*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, second edn, John Wiley & Sons Inc., New York. A Wiley-Interscience Publication.

Jung, R. & Tremayne, A. (2011), 'Useful models for time series of counts or simply wrong ones?', *AStA Advances in Statistical Analysis* **95**, 59–91.

Kedem, B. & Fokianos, K. (2002), *Regression models for time series analysis*, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ.

Lindley, D. V. (1956), the estimation of velocity dis tributions from counts, *in* 'Proceedings of the Internationl Congress of Mathematicians', Vol. 3, North-Holland, Amsterdam, pp. 427–444.

Mátyás, L., ed. (1999), *Generalized method of moments estimation*, Cambridge University Press, Cambridge.

McKenzie, E. (2003), Discrete variate time series, *in* 'Stochastic processes: modelling and simulation', Vol. 21 of *Handbook of Statistics*, North-Holland, Amsterdam, pp. 573–606.

Rajput, B. & Rosinski, J. (1989), 'Spectral representation of infinitely divisible distributions', *Probability Theory and Related Fields* **82**, 451–487.

Reynolds, J. F. (1968), 'On the autocorrelation and spectral functions of queues', *Journal of Applied Probability* **5**, 467–475.

Sato, K. (1999), *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge.

Shephard, N., ed. (2005), *Stochastic Volatility: Selected Readings*, Oxford University Press, Oxford.

Steutel, F. W. & van Harn, K. (1979), 'Discrete analogues of self-decomposability and stability', *The Annals of Probability* **7**(5), 893–899.

Surgailis, D., Rosinski, J., Mandrekar, V. & Cambanis, S. (1993), 'Stable mixed moving averages', *Probability Theory and Related Fields* **97**, 543–558.

Veraart, A. & Winkel, M. (2010), Time change, *in* R. Cont, ed., 'Encyclopedia of Quantitative Finance', Vol. 4, Wiley, pp. 1812–1816.

Weiß, C. (2008), 'Thinning operations for modeling time series of counts: a survey', *AStA Advances in Statistical Analysis* **92**, 319–341.

Wolpert, R. L. & Brown, L. D. (2011), Stationary infinitely-divisible Markov processes with non-negative integer values. Working paper, April 2011.

Wolpert, R. L. & Taqqu, M. S. (2005), 'Fractional Ornstein–Uhlenbeck Lévy processes and the Telecom process: Upstairs and downstairs', *Signal Processing* **85**, 1523–1545.

Zhu, R. & Joe, H. (2003), 'A new type of discrete self-decomposability and its application to continuous-time Markov processes for modeling count data time series', *Stochastic Models* **19**(2), 235–254.