Modeling, simulation and inference for multivariate time series of counts using trawl processes

Almut E. D. Veraart
Imperial College London

2018 Workshop on Finance, Insurance, Probability and Statistics (FIPS 2018)
King's College London, 10-11 September 2018

Introduction

Aim of the Project

- Modelling **multivariate time series of counts**.
- Count data: Non-negative and integer-valued, and often over-dispersed (i.e. variance $>$ mean).
- Various applications: Medical science, epidemiology, meteorology, network modelling, actuarial science, econometrics and finance.

- Modelling **multivariate time series of counts**.
- Count data: Non-negative and integer-valued, and often over-dispersed (i.e. variance $>$ mean).
- Various applications: Medical science, epidemiology, meteorology, network modelling, actuarial science, econometrics and finance.

Aim of the project

Develop continuous-time models for time series of counts that

- allows for a flexible autocorrelation structure;
- can deal with a variety of marginal distributions;
- allows for flexibility when modelling cross-correlations;
- is analytically tractable.

- Overall, two predominant modelling approaches:
  - Discrete autoregressive moving-average (DARMA) models introduced by Jacobs & Lewis (1978a,b).

    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 (see McKenzie (2003)).
  - Models obtained from thinning operations going back to the influential work of Steutel & van Harn (1979), e.g. INARMA. See also Zhu & Joe (2003) for related more recent work.

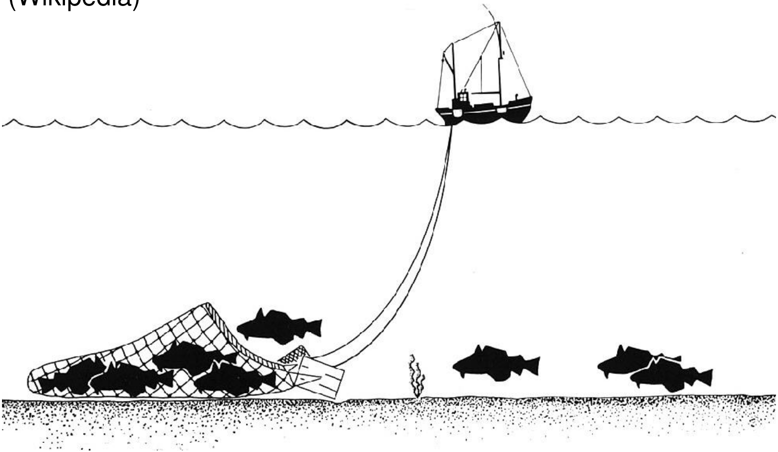
- Overall, two predominant modelling approaches:
  - Discrete autoregressive moving-average (DARMA) models introduced by Jacobs & Lewis (1978a,b).

    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 (see McKenzie (2003)).
  - Models obtained from thinning operations going back to the influential work of Steutel & van Harn (1979), e.g. INARMA. See also Zhu & Joe (2003) for related more recent work.
- Key idea of this paper: Use trawling for modelling counts! – Nested within the framework of ambit fields (Barndorff-Nielsen & Schmiegel (2007)) and extends results by Barndorff-Nielsen, Pollard & Shephard (2012) and Barndorff-Nielsen, Lunde, Shephard & Veraart (2014).

Introduction

What is trawling...? A first "definition"

"Trawling is a method of fishing that involves pulling a fishing net through the water behind one or more boats. The net that is used for trawling is called a trawl." (Wikipedia)



- Let N be a homogeneous Poisson random measure on $\mathbb{R}^n \times \mathbb{R}^2$ with compensator
$$\mathbb{E}(N(d\mathbf{y}, dx, dt)) = \nu(d\mathbf{y})dxdt,$$
where ν is a Lévy measure satisfying $\int_{-\infty}^{\infty} \min(1, ||\mathbf{y}||)\nu(d\mathbf{y}) < \infty$.
- Assume that N is positive integer-valued, i.e. ν is concentrated on $\mathbb{N}$.
- Then we define an $\mathbb{N}^n$-valued, homogeneous Lévy basis on $\mathbb{R}^2$ in terms of the Poisson random measure as
$$\mathbf{L}(dx, dt) = (L^{(1)}(dx, ds), \ldots, L^{(n)}(dx, ds))' = \int_{-\infty}^{\infty} \mathbf{y}N(d\mathbf{y}, dx, dt). \quad (1)$$
- The Lévy basis $\mathbf{L}$ is infinitely divisible with cumulant function
$$C_{\mathbf{L}(dx,dt)}(\theta) := \log(\mathbb{E}(\exp(i\theta L(dx, dt)))) = \int_{\mathbb{R}} \left(e^{i\theta\mathbf{y}} - 1\right) \nu(d\mathbf{y})dxdt$$
$$= C_{\mathbf{L}'}(\theta)dxdt, \text{ where } \mathbf{L}' \text{ is the Lévy seed.}$$

Theoretical framework

Integer-valued, homogeneous Lévy bases: The cross-correlation

- From Feller (1968), we know that any non-degenerate distribution on $\mathbb{N}^n$ is infinitely divisible if and only if it can be expressed as a discrete compound Poisson distribution.
- A random vector with infinitely divisible distribution on $\mathbb{N}^n$ always has non-negatively correlated components.
- We model the Lévy seed by an n-dimensional compound Poisson process given by

$$\mathbf{L}'_t = \sum_{j=1}^{N_t} \mathbf{Z}_j,$$

where $N = (N_t)_{t \geq 0}$ is an homogeneous Poisson process of rate $v > 0$ and the $(\mathbf{Z}_j)_{j \in \mathbb{N}}$ form a sequence of i.i.d. random variables independent of N and which have no atom in $\mathbf{0}$, i.e. not all components are simultaneously equal to zero, more precisely, $\mathbb{P}(\mathbf{Z}_j = \mathbf{0}) = 0$ for all j.

Definition 1

A trawl for the ith component is a Borel set $A^{(i)} \subset \mathbb{R} \times (-\infty, 0]$ such that $Leb(A^{(i)}) < \infty$. Then, we set

$$A_t^{(i)} = A^{(i)} + (0, t), \qquad i \in \{1, \ldots, n\}.$$

Definition 1

A trawl for the ith component is a Borel set $A^{(i)} \subset \mathbb{R} \times (-\infty, 0]$ such that $Leb(A^{(i)}) < \infty$. Then, we set

$$A_t^{(i)} = A^{(i)} + (0, t), \qquad i \in \{1, \ldots, n\}.$$

- Typically, we choose $A^{(i)}$ to be of the form

$$A^{(i)} = \{(x, s) : s \leq 0,\ 0 \leq x \leq d^{(i)}(s)\}, \tag{2}$$

  where $d^{(i)} : (-\infty, 0] \mapsto \mathbb{R}$ is a cont. and $Leb(A^{(i)}) < \infty$.
- Then $\quad A_t^{(i)} = A^{(i)} + (0, t) = \{(x, s) : s \leq t,\ 0 \leq x \leq d^{(i)}(s - t)\}$.
- If $d^{(i)}$ is also monotonically non-decreasing, then $A^{(i)}$ is a *monotonic trawl*.

Definition 2

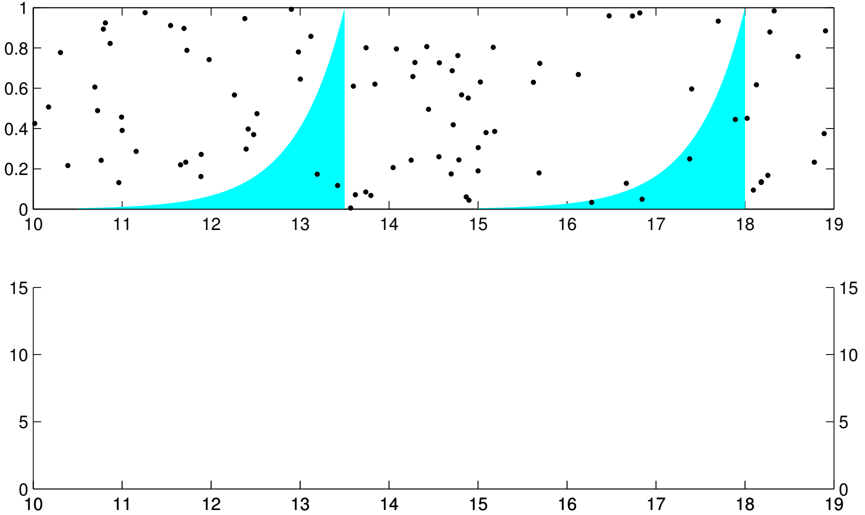
We define an *n-dimensional stationary integer-valued trawl (IVT) process* $(\mathbf{Y}_t)_{t\geq 0}$ by $\mathbf{Y}_t = (L^{(1)}(A_t^{(1)}), \ldots, L^{(n)}(A_t^{(n)}))'$,

- where
$$L^{(i)}(A_t^{(i)}) = \int_{\mathbb{R}\times\mathbb{R}} \mathbf{I}_{A^{(i)}}(x, s-t) L^{(i)}(dx, ds), \quad i \in \{1, \ldots, n\}.$$
- $\mathbf{L}$ is the n-dimensional integer-valued, homogeneous Lévy basis on $\mathbb{R}^2$ (see (1)).
- $A_t^{(i)} = A^{(i)} + (0, t)$ with $A^{(i)} \subset \mathbb{R} \times (-\infty, 0]$ and $Leb(A^{(i)}) < \infty$ is the trawl.

- Wolpert & Taqqu (2005) study a subclass of general (univariate) trawl processes (not necessarily restricted to IV) under the name "up-stairs" representation, "random measure of a moving geometric figure in a higher-dimensional space"
- Wolpert & Brown (2011) study so-called "random measures processes" which also fall into the (univariate) trawling framework.

Theoretical framework

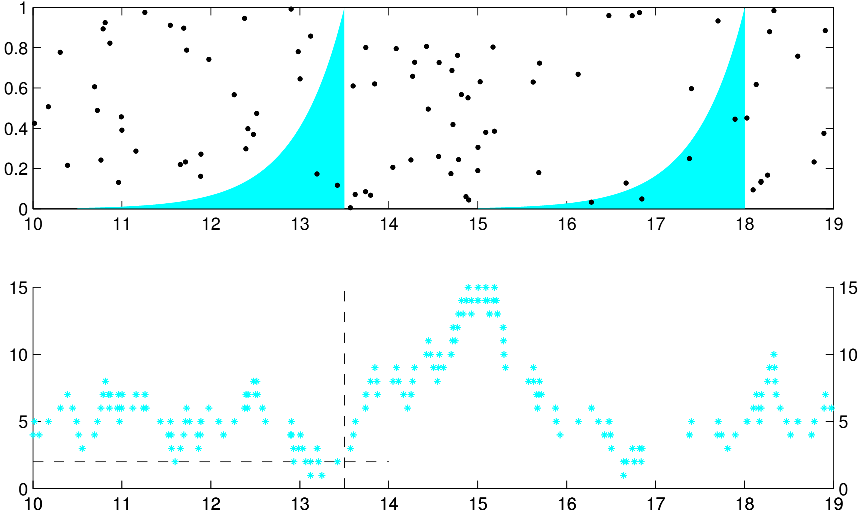
Definition of a Trawl Process



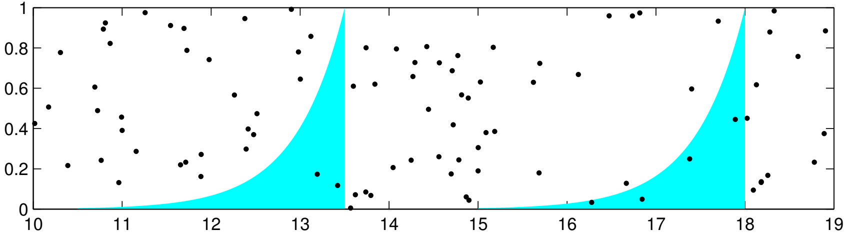
Imperial College London

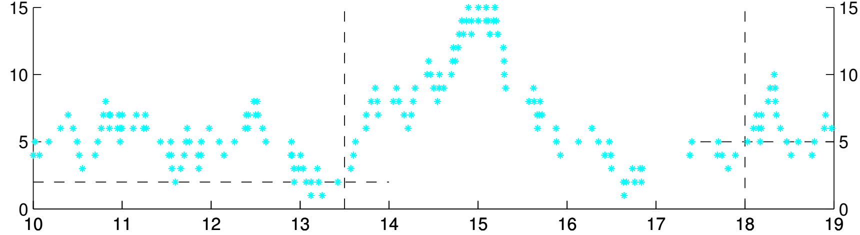
Theoretical framework

Definition of a Trawl Process



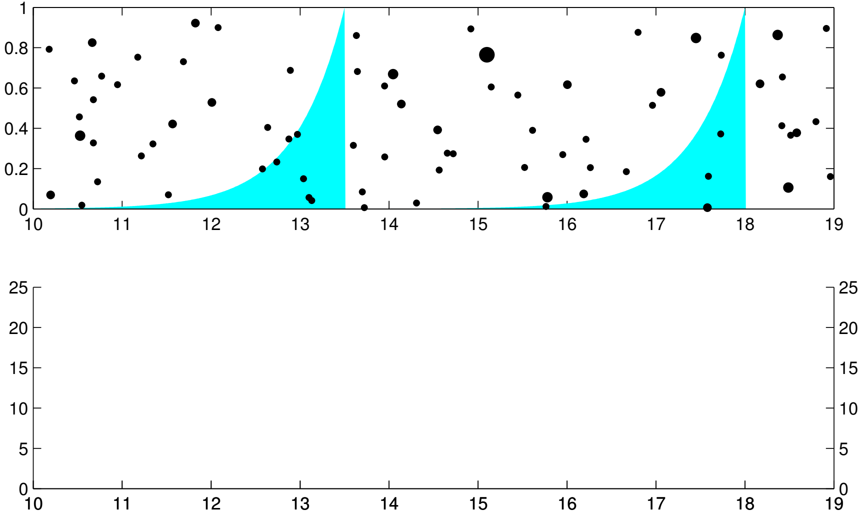
Theoretical framework
Definition of a Trawl Process





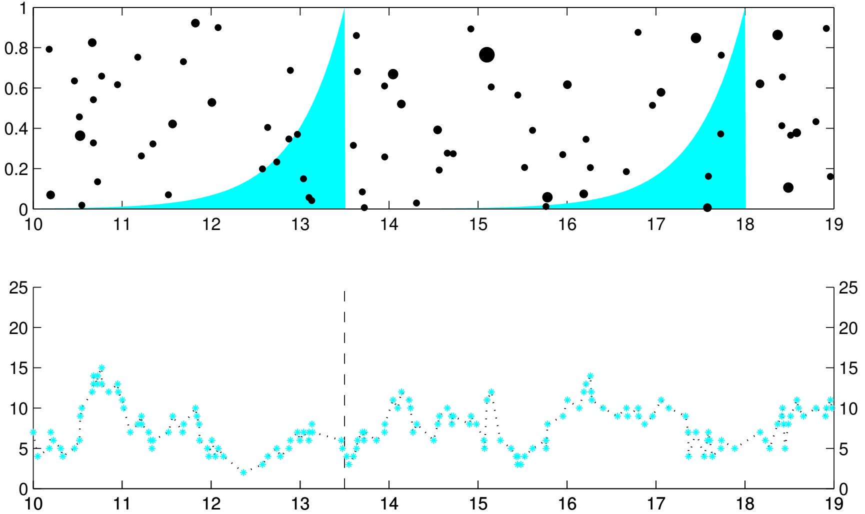
Examples

Negative Binomial exponential-trawl process



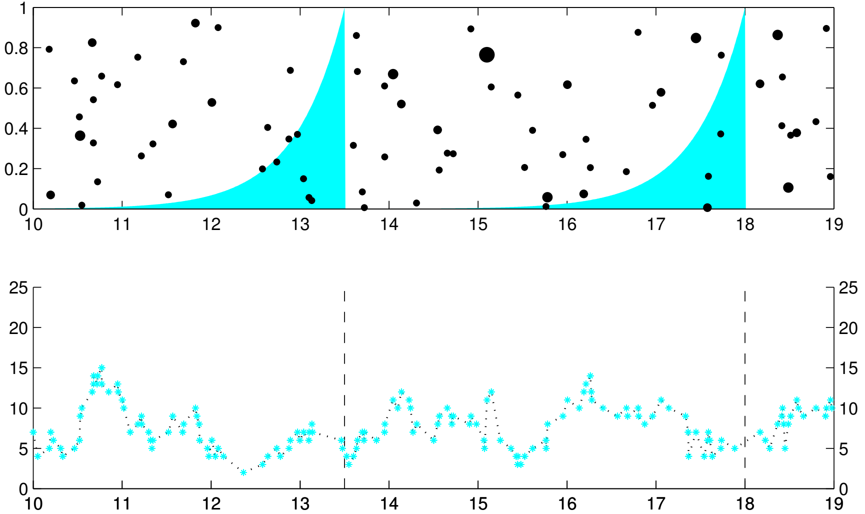
Examples

Negative Binomial exponential-trawl process



Examples

Negative Binomial exponential-trawl process



- The IVT process is stationary and infinitely divisible.
- The IVT process is mixing $\Rightarrow$ weakly mixing $\Rightarrow$ ergodic.
- The cumulant function of a trawl process is given by

$$C_{Y_t^{(i)}}(\theta) = C_{L^{(i)}(A_t^{(i)})}(\theta) = Leb(A^{(i)}) C_{L'^{(i)}}(\theta),$$

  - I.e. 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.

- The covariance between two (possibly shifted) components $1 \leq i \leq j \leq n$ for $t, h \geq 0$ is given by

$$\mathrm{Cov}\left(L^{(i)}(A_t^{(i)}), L^{(j)}(A_{t+h}^{(j)})\right) = Leb\left(A^{(i)} \cap A_h^{(j)}\right)\left(\int_{\mathbb{R}}\int_{\mathbb{R}} y_i y_j \nu^{(i,j)}(dy_i, dy_j)\right),$$

$$Cor(L^{(i)}(A_t^{(i)}), L^{(i)}(A_{t+h}^{(i)})) = \frac{Leb(A^{(i)} \cap A_h^{(i)})}{Leb(A^{(i)})}.$$

Multivariate law of the Lévy seed

Poisson mixtures

- The law of $\mathbf{L}'$ is of discrete compound Poisson type by construction.
- Use Poisson mixtures based on random additive effect models, see Barndorff-Nielsen et al. (1992).
- Consider random variables $X_1, \ldots, X_n$ and $Z_1, \ldots, Z_n$, such that, conditionally on $\{Z_1, \ldots, Z_n\}$, the $X_1, \ldots, X_n$ are independent and Poisson distributed with means given by the $\{Z_1, \ldots, Z_n\}$.
- Model the joint distribution of the $\{Z_1, \ldots, Z_n\}$ by a so-called additive effect model as follows:

$$Z_i = \alpha_i U + V_i, \quad i = 1, \ldots, n,$$

  where the random variables $U, V_1, \ldots, V_n$ are independent and the $\alpha_1, \ldots, \alpha_n$ are nonnegative parameters.
- We have explicit formulas for the joint law of $(X_1, \ldots, X_n)$ and

$$\mathbb{E}(X_i) = \alpha_i \mathbb{E}(U) + \mathbb{E}(V_i), \quad i = 1, \ldots, n,$$
$$\mathrm{Cov}(X_i, X_j) = \alpha_i \alpha_j \mathrm{Var}(U), \text{ if } i \neq j.$$

Theoretical results

Representation as compound Poisson distribution and as a multivariate negative binomial distribution

- The Poisson mixture model of random-additive-effect type can be represented as a compound Poisson distribution
- If U and V_is follow suitable Gamma distributions, then negative binomial marginal law can be achieved allowing for
  - 1) independence,
  - 2) complete dependence, or
  - 3) dependence with additional independent factors between the components.
- Recall that the (multivariate) negative binomial distribution can be represented as a compound Poisson distribution with (multivariate) logarithmic jump size distribution.
- The representation result is used in the simulation algorithm for the multivariate trawl process.

Key properties of IVT processes

Overview

Flexible marginal distributions:

- Poisson trawl process;
- Negative binomial trawl process;
- other compound Poisson distributions.

Key properties of IVT processes

Overview

Flexible marginal distributions:

- Poisson trawl process;
- Negative binomial trawl process;
- other compound Poisson distributions.

Various choices of the trawl function:

- Superpositions of exponential trawls: $d^{(i)}(z) = \int_0^\infty e^{\lambda z} \pi^{(i)}(d\lambda)$, for $z \leq 0$, for a probability measure $\pi^{(i)}$ on $(0, \infty)$.
- Possibility of allowing for long memory.

Inference

(Generalised) Method of Moments

- Use a (generalised) method of moments in a two-stage equation-by-equation approach to estimate the marginal parameters first, followed by the dependence parameters.

- Step 1a) Use the acf $r^{(i)}(h) = \frac{Leb(A^{(i)} \cap A_h^{(i)})}{Leb(A^{(i)})}$ to infer the trawl parameters.

- Step 1b) Use the cumulant function $C_{Y_t^{(i)}}(\theta) = Leb(A^{(i)}) C_{L'^{(i)}}(\theta)$ to infer the marginal parameters of the Lévy basis.

- Step 2a) Compute $Leb(A^{(i)} \cap A^{(j)})$ for $i \neq j$.

- Step 2b) Use the cross-covariance function

$$Cov\left(L^{(i)}(A_t^{(i)}), L^{(j)}(A_t^{(j)})\right) = Leb\left(A^{(i)} \cap A^{(j)}\right) \left(\int_{\mathbb{R}} \int_{\mathbb{R}} y_i y_j \nu^{(i,j)}(dy_i, dy_j)\right)$$

to infer the dependence parameters.

Imperial College London

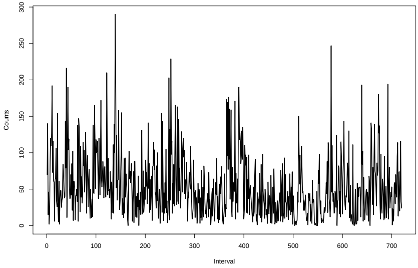
Empirical illustration
High frequency financial data from LOBSTER

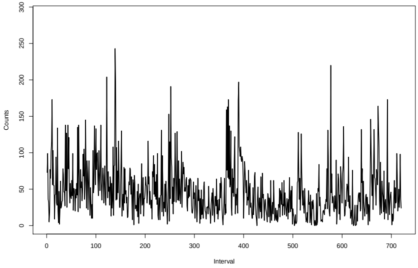
- Study high frequency *limit order book data* from LOBSTER.
- We picked the Apple data for August 8, 2017: Start at 11:00am, end at 12:00 (noon).
- We analyse the joint behaviour of the number of newly submitted and fully deleted limit orders over 5s intervals (720 observations in total).
- We fit a bivariate trawl model with double exponential trawl function and bivariate negative binomial law.
- Summary statistic:

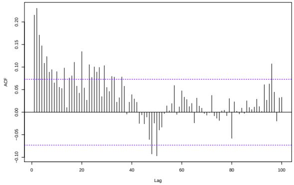
	Min	1st Quartile	Median	Mean	3rd Quartile	Max
No. of submissions	0.00	19.00	44.00	53.46	74.25	290.00
No. of cancellations	0.00	22.00	38.00	46.77	63.00	243.00

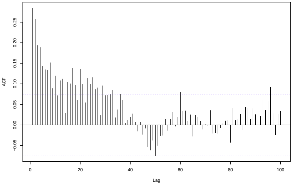
Number of limit order submissions and deletions

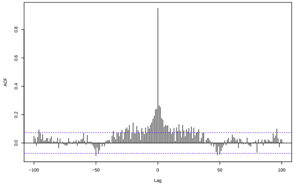
Time series, acf and crosscorrelation





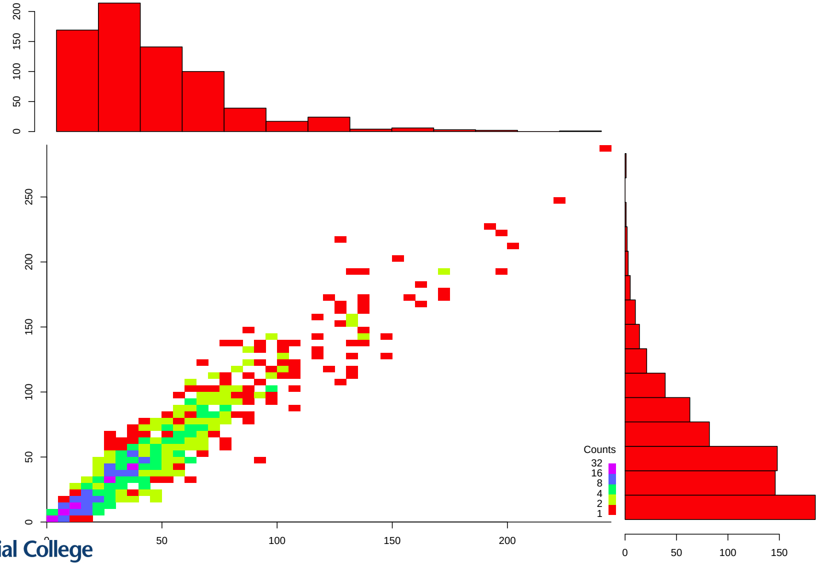






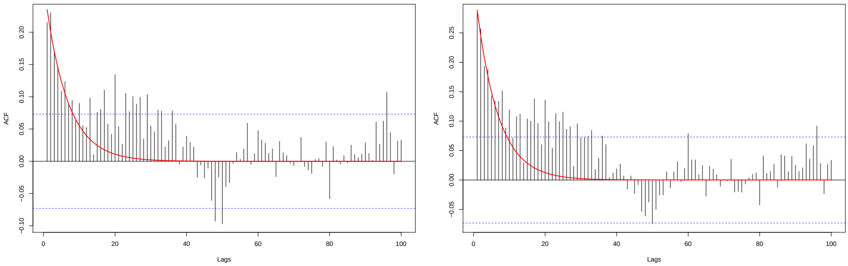
Number of limit order submissions and deletions

Histograms: submissions (right), full cancellations (top)



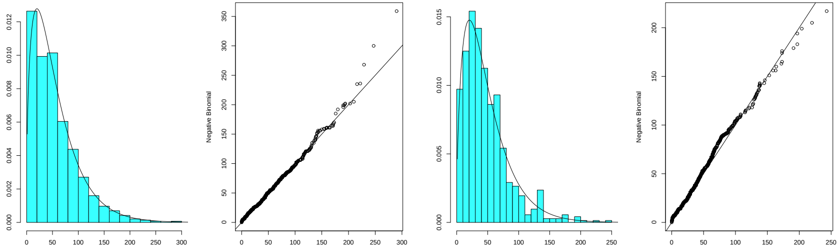
Empirical illustration

Fitted trawl (sum of two exponentials) to number of submissions and deletions



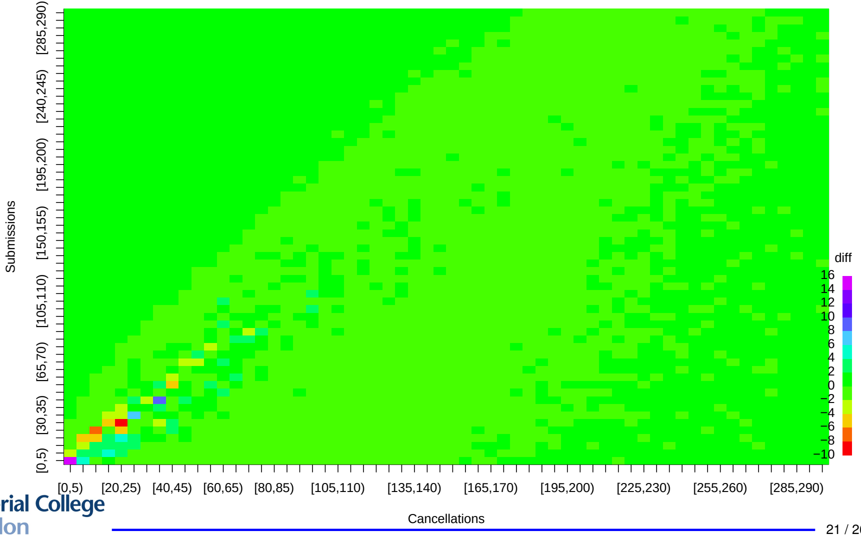
Empirical illustration

Fitted bivariate negative binomial marginal fit (to number of submissions and deletions)



Empirical illustration

Fitted bivariate negative binomial: Bivariate histogram assessment

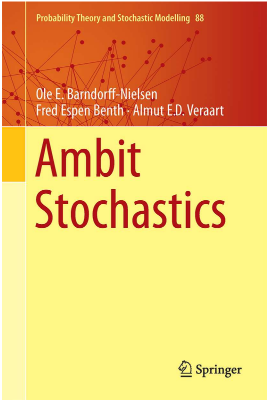


Main contributions

- New continuous-time framework for modelling multivariate stationary, serially correlated count data.
- Two key components:
  - Integer-valued, homogeneous Lévy basis: Generates random point pattern and determines marginal distribution and cross-sectional dependence.
  - Trawl: Thins the point pattern and determines the autocorrelation structure.
- Simulation algorithm & parameter estimation of IVT processes with monotonic trawl.
- Simulation studies reveal good performance of (generalised) method of moments and quasi-maximum-likelihood methods for parameter estimation.
- Empirical application to high frequency financial data.

You want to know more...?!

- Preprint: Available on my website (article forthcoming in the Journal of Multivariate Analysis).
- New R package `trawl` available on CRAN.
- Many more details in our new book:



Imperial College London

Barndorff-Nielsen, O. E., Blæsild, P., & Seshadri, V. (1992). Multivariate distributions with generalized inverse Gaussian marginals, and associated Poisson mixtures. *The Canadian Journal of Statistics. La Revue Canadienne de Statistique*, 20(2), 109–120.

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

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* (pp. 93–124). Heidelberg: Springer.

Bibliography II

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

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

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

Veraart, A. E. D. (2015). Modelling multivariate serially correlated count data in continuous time. In preparation.

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

Bibliography III

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.