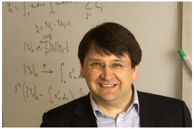
# Likelihood-based estimation, model selection, and forecasting of integer-valued trawl processes

Almut E. D. Veraart
Imperial College London

New Results on Time Series and their Statistical Applications
CIRM Luminy, 14-18 September 2020

# Collaborators

This is joint work with

- Mikkel Bennedsen (Aarhus University) 
- Asger Lunde (Aarhus University) 
- Neil Shephard (Harvard University) 

# Introduction

- **Time series of counts** appear in various applications: Medical science, epidemiology, meteorology, network modelling, actuarial science, econometrics and finance.
- Count data: Non-negative and integer-valued, and often over-dispersed (i.e. variance $>$ mean).
- Recently the class of (integer-valued) trawl (IVT) processes has been introduced as a flexible model, see Barndorff-Nielsen et al. (2014) for the univariate and Veraart (2019) for the multivariate case.

## Aim of the project

- Improve the estimation method for IVT processes (likelihood-based rather than moment-based);
- Tailor model selection tools to the IVT class;
- (Probabilistic) forecasting of IVT processes;

# A *very* short and incomplete review of the literature

- Recent surveys & some new developments:
  Cameron & Trivedi (1998), Kedem & Fokianos (2002),Cui & Lund (2009); Davis et al. (1999); Davis & Wu (2009); Jung & Tremayne (2011); McKenzie (2003); Weiß (2008), Karlis (2016), Fokianos (2016).
- Literature on count data is spread across different disciplines.
- Overall, two predominant modelling approaches:
  - Discrete autoregressive moving-average (DARMA) models introduced by Jacobs & Lewis (1978a,b).
  - Models obtained from thinning operations going back to the influential work of Steutel & van Harn (1979), e.g. INAR(MA), see e.g. Pedeli et al. (2015).
- Further models: Regression type models (typically based on generalised linear models, see e.g. Fokianos (2016)), also Fokianos et al. (2020); state-space and Bayesian approaches.

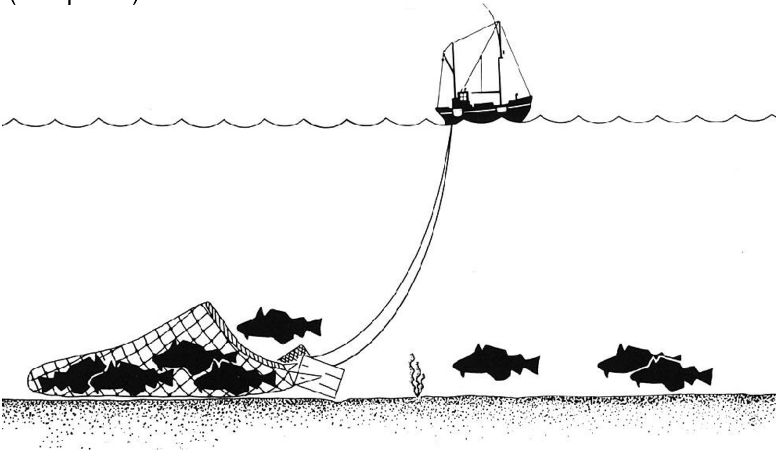
## Our approach:

- Use ”trawling” for modelling counts.
- This is a continuous-time framework based on the idea of ”thinning” points.

# 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)



We define a stationary integer-valued trawl (IVT) process $(X_t)_{t\geq 0}$ by

$$X_t = L(A_t) = \int_{\mathbb{R}\times\mathbb{R}} \mathbf{I}_A(x, s-t) L(dx, ds).$$

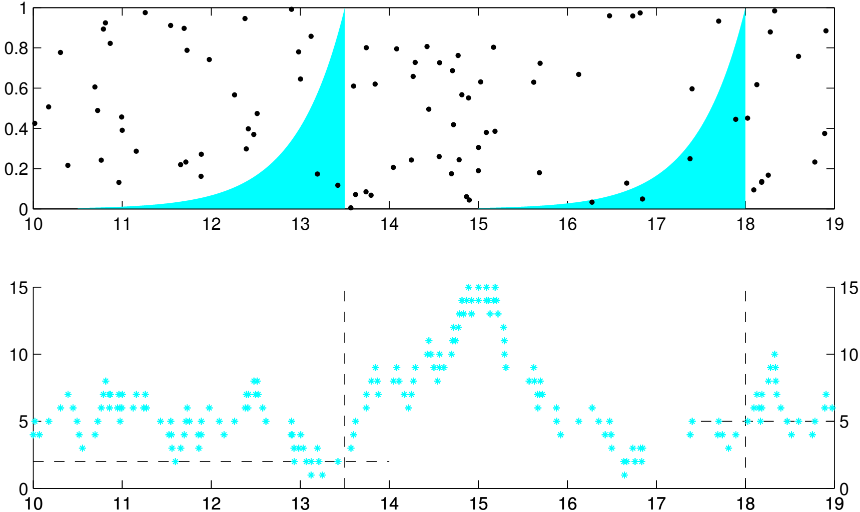
- $L$ is the integer-valued, homogeneous Lévy basis on $[0, 1] \times \mathbb{R}$:
  - $L(dx, ds) := \int_{-\infty}^{\infty} yN(dy, dx, ds), \quad (x, s) \in [0, 1] \times \mathbb{R}$.
  - $N$ is a homogeneous Poisson random measure on $\mathbb{Z} \times [0, 1] \times \mathbb{R}$ with compensator $\eta \otimes \mathrm{Leb} \otimes \mathrm{Leb}$, i.e. $\mathbb{E}(N(dy, dx, ds)) = \eta(dy)dxds$, where $\eta$ is a Lévy measure satisfying $\int_{-\infty}^{\infty} \min(1, |y|)\eta(dy) < \infty$.
- A Borel set $A_t = A + (0, t)$ with $A = A_0 \subseteq [0, 1] \times (-\infty, 0]$ and $\mathrm{Leb}(A) < \infty$ is called the trawl.
  - Typically, we choose $A$ to be of the form

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

where $d : (-\infty, 0] \mapsto [0, 1]$ is continuous and $\mathrm{Leb}(A) < \infty$.

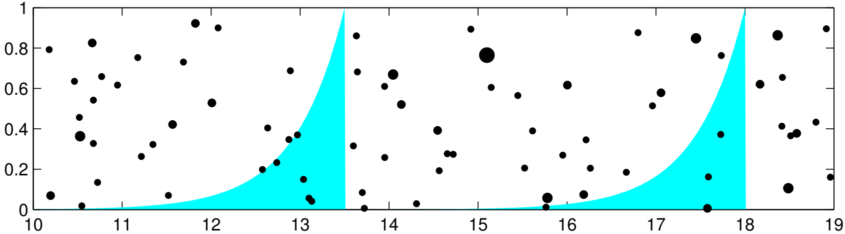
# Example

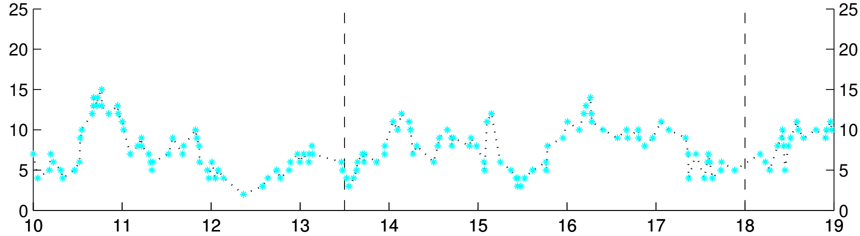
## Poisson-Exponential trawl



# Example

## Negative binomial-Exponential trawl



# Some key properties of IVT processes
## Cumulants

- The IVT process is stationary and infinitely divisible.
- The IVT process is mixing $\Rightarrow$ weakly mixing $\Rightarrow$ ergodic.
- The cumulant (log-characteristic) function of a trawl process is, for $\theta \in \mathbb{R}$, given by

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

  where the random variable $L'$ (called the Lévy seed) associated with $L$ satisfies

$$\mathbb{E}[\exp(i\theta L')] = \exp(C_{L'}(\theta)), \quad \text{with} \quad C_{L'}(\theta) = \int \left(e^{i\theta y} - 1\right) \eta(dy).$$

  - I.e. to any infinitely divisible integer-valued law $\pi$, say, there exists a stationary integer-valued trawl process having $\pi$ as its one-dimensional marginal law.
- The autocorrelation function is given by

$$\rho(h) := \mathrm{Cor}(Y_t, Y_{t+h}) = \frac{\mathrm{Leb}(A \cap A_h)}{\mathrm{Leb}(A)}, \qquad \text{for } h > 0.$$

### Example 1 (Poissonian Lévy seed)

Let $L' \sim$ Poisson$(\nu)$. Then $X_t \sim$ Poisson$(\nu \mathrm{Leb}(A))$, i.e., for all $t \geq 0$,

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

### Example 2 (Negative Binomial Lévy seed)

Let $L' \sim$ NB$(m, p)$ for $m > 0, p \in [0, 1]$. Then $X_t \sim$ NB$(m\mathrm{Leb}(A), p)$, i.e., for all $t \geq 0$,

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

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

- Recall the typical choice for the trawl:

$$A = A_0 = \{(x, s) : s \leq 0,\ 0 \leq x \leq d(s)\}, \qquad A_t = A + (0, t).$$

- Restrict attention to a class of *superposition trawls*:

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

  where $\pi$ is a probability measure on $\mathbb{R}_+$.
- For $h \geq 0$, the acf is given by

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

- **Exponential trawl function:** Let $\lambda > 0$ and $\pi(dx) = \delta_\lambda(dx)$, then $d(s) = e^{\lambda s}$ for $s \leq 0$ and

$$\rho(h) = \mathrm{Cor}(X_{t+h}, X_t) = \exp(-\lambda h), \quad h \geq 0.$$

- **Inverse Gaussian trawl function:** Letting $\pi$ be given by the inverse Gaussian distribution

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

where $K_\nu(\cdot)$ is the modified Bessel function of the third kind and $\gamma, \delta \geq 0$ with both not zero simultaneously. Then

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

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

# Examples

## Modelling the trawl function/correlation structure

- **Gamma trawl function:** Let $\pi$ have the $\Gamma(1+H,\alpha)$ density,

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

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

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

and

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

Note that in this case

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

i.e. the trawl process has long memory for $H \in (0,1]$.

# Estimation

## From method of moments to composite likelihood

- Suppose we have $n \in \mathbb{N}$ observations of the IVT process $X$, $x_1, \ldots, x_n$, on an equidistant grid of size $\Delta = T/n$.
- Define

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

- Let $\Theta$ be a compact parameter space such that the true parameter vector, $\theta_0$, lies in the interior of $\Theta$.
- Construct the composite likelihood function, for $\mathcal{H} \subseteq \{1, 2, \ldots, n-1\}$,

$$L_{CL}^{\mathcal{H}}(\theta; x) := \prod_{h \in \mathcal{H}} CL^{(h)}(\theta; x) = \prod_{h \in \mathcal{H}} \prod_{i=1}^{n-h} f(x_{i+h}, x_i; \theta).$$

- The maximum composite likelihood (MCL) estimator of $\theta$ is defined as

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

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

# Pairwise likelihood

## The general case and a simulation-based approach

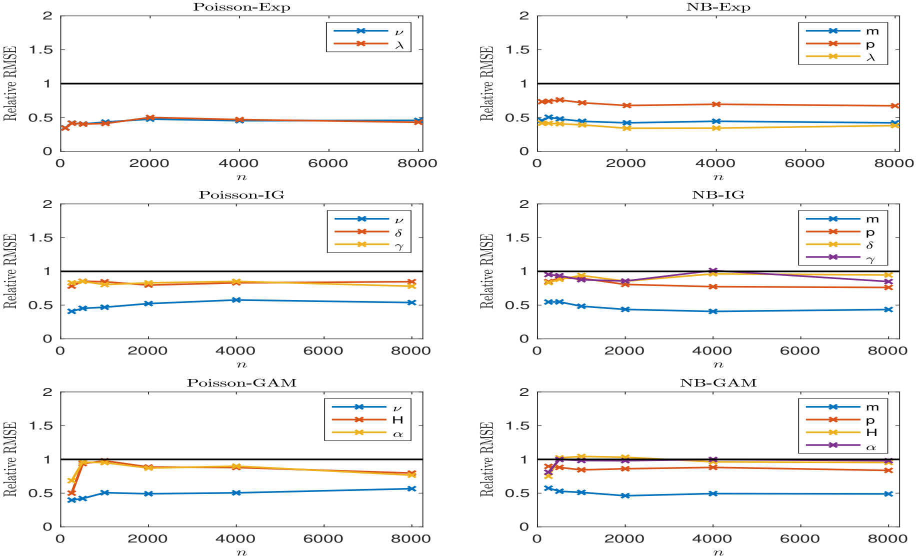
- The joint probability mass function of two observations $x_{i+h}$ and $x_i$ is

$$f(x_{i+h},x_i;\theta) := P_\theta\left(X_{(i+h)\Delta} = x_{i+h}, X_{i\Delta} = x_i\right)$$
$$= \sum_{c=-\infty}^{\infty} P_\theta\left(L(A_{(i+h)\Delta} \setminus A_{i\Delta}) = x_{i+h} - c\right)$$
$$\cdot P_\theta\left(L(A_{i\Delta} \setminus A_{(i+h)\Delta}) = x_i - c\right)$$
$$\cdot P_\theta\left(L(A_{(i+h)\Delta} \cap A_{i\Delta}) = c\right).$$

- Suppose the Lévy basis $L$ is positive, i.e. $\eta(y) = 0$ for $y < 0$. Then we can replace $\sum_{c=-\infty}^{\infty}$ by $\sum_{c=0}^{\min\{x_{i+h},x_i\}}$ in the above formula.
- Let $t, s \geq 0$, choose $C \in \mathbb{N}$ and let $c^{(j)} \sim L(A_t \cap A_s)$, $j = 1, 2, \ldots, C$, be an iid sample. A simulation based unbiased estimator of $f(x_t, x_s; \theta)$ is

$$\hat{f}(x_t, x_s; \theta) = \frac{1}{C}\sum_{j=1}^{C} P_\theta(L(A_t \setminus A_s) = x_t - c^{(j)})P_\theta(L(A_s \setminus A_t) = x_s - c^{(j)}).$$

# MCL outperforms GMM for IVTs




RMSE of the MCL estimator divided by the RMSE of the GMM estimator.

# Model selection for IVTs

- Following Takeuchi (1976), Varin & Vidoni (2005) we apply the composite likelihood information criterion (CLAIC)

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

  as a basis for model selection.

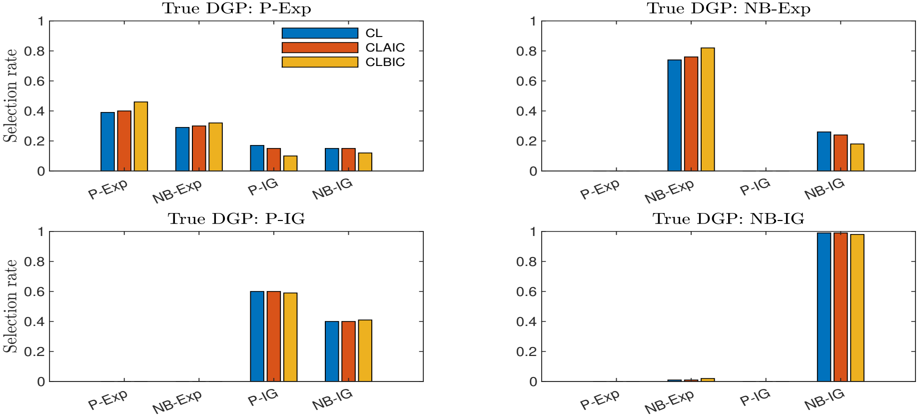
- Note that $G(\theta)^{-1} = H(\theta)^{-1}V(\theta)H(\theta)^{-1}$ is the asymptotic covariance matrix of the MCE. We use the straight-forward estimator $\hat{H}(\hat{\theta}^{CL}) = -n^{-1}\frac{\partial}{\partial\theta\partial\theta'}l_{CL}(\hat{\theta}^{CL}; x)$ which is consistent for $H(\theta)$ due to the stationarity and ergodicity of the IVT process, and estimate $\hat{V}(\hat{\theta}^{CL})$ by parametric bootstrap.

- We also apply the BIC adapated to the composite likelihood case, see Gao & Song (2010)

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

  where $n$ is the number of observations of the data series $x$.

# Simulation study of model selection procedure




The numbers plotted are average selection rates of the models given on the $x$-axis, using a given criteria over $M = 100$ Monte Carlo simulations. For each Monte Carlo replication, $n = 4\,000$ observations of the true DGP are simulated on a grid with step size $\Delta = 0.1$.

# Probabilistic forecasting of IVTs

- Let $\mathcal{F}_t = \sigma((X_s)_{s \leq t})$, let $h > 0$ be a forecast horizon.
- Goal: Forecast the future value $X_{t+h}$ (and its distribution).
- Note that $\tilde{X}_{t+h|t} = \mathbb{E}[X_{t+h}|\mathcal{F}_t]$ is not *data coherent*.
- Consider instead a probabilistic forecasting approach, where the interest is in the distribution of $X_{t+h}|\mathcal{F}_t$ and generate data coherent point forecasts, e.g. using the median or mode of the distribution.
- However, since the IVT process $X_t$ is in general non-Markovian, the distribution of $X_{t+h}|\mathcal{F}_t$ is highly intractable.
- The probabilistic forecast of $X_{t+h}|X_t$ gives promising results.

# Probabilistic forecasting of IVTs

## Proposition 1

*Suppose the Lévy basis* $L$ *is positive, i.e.* $\eta(y) = 0$ *for* $y < 0$*. Now*

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

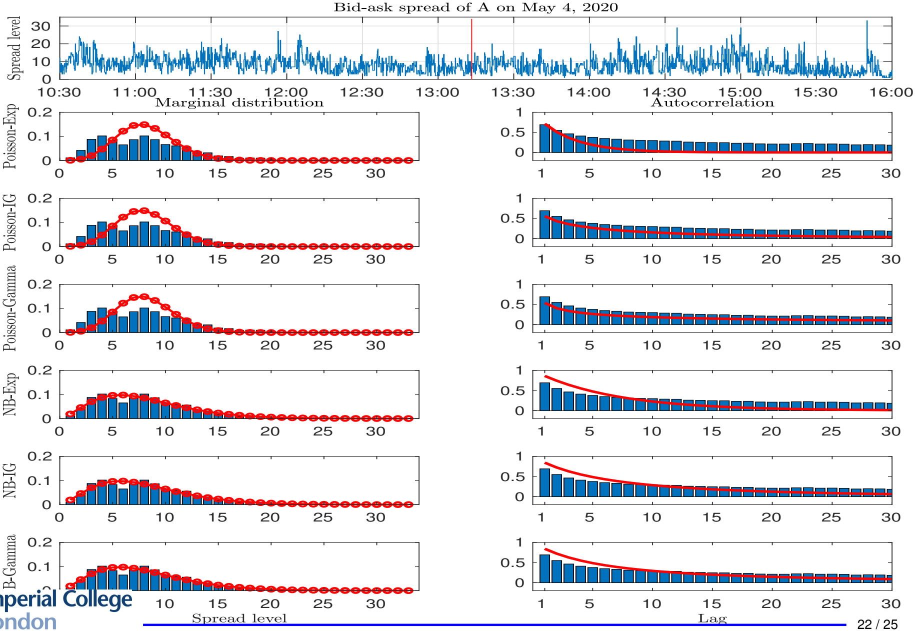
*where*

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

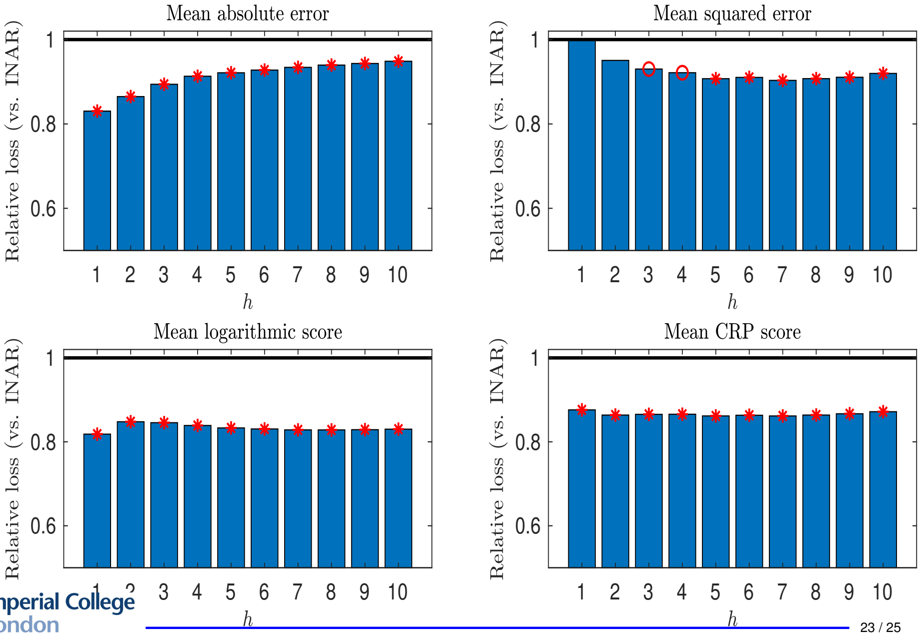
# Empirical study: Goal: Forecasting the bid-ask spread

- Study of high frequency data of bid-ask spreads of equity prices.
- Spread data of Agilent Technologies Inc. stock (ticker: A, NYSE) (measured in U.S. dollar cents); single day, May 4, 2020 [used data from 10:30AM to 4PM, i.e. discarded the first 60 minutes]
- The data were cleaned using the approach proposed in Barndorff-Nielsen et al. (2009), then sampled equidistantly (5s) using the previous-tick method resulting in $n = 3\,961$ observations.
- Let $s_t$ be the spread level at time $t$. Since the minimum spread level in the data is one tick (i.e. one cent), we work on $x_t = s_t - 1$.
- Model selection: The NB-Gamma model is the preferred model on all three criteria, while the second-best model is the NB-IG model. Note, however, that these two models appear to provide an almost identical fit.

# Empirical study: Model selection



# Empirical study: Forecasting the bid-ask spread



# Summary

- Integer-valued trawl processes provide a continuous-time framework for modelling stationary, serially correlated count data.
- They consist of two key components:
  - Integer-valued, homogeneous Lévy basis: Generates random point pattern and determines marginal distribution.
  - Trawl: Thins the point pattern and determines the autocorrelation structure.
- We showed that the pairwise likelihood for IVTs is tractable and MCL outperforms previously used GMM.
- The pairwise likelihood can be used for model selection criteria (CLBIC slightly preferred)
- Method for probabilistic forecasting of IVTs using the pairwise likelihood principle.
- Application to forecasting equity spread data: Superior performance of IVT compared to INAR(1) benchmark model.

# Bibliography I


Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: trades and quotes. *The Econometrics Journal*, 12(3), C1–C32. 21

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. 3

Bennedsen, M., Lunde, A., Shephard, N., & Veraart, A. E. D. (2020). Likelihood-based estimation, model selection, and forecasting of integer-valued trawl processes. Work in progress.

Cameron, A. C. & Trivedi, P. K. (1998). *Regression analysis of count data*, volume 30 of *Econometric Society Monographs*. Cambridge: Cambridge University Press. 4

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

Davis, R. A., Wang, Y., & Dunsmuir, W. T. M. (1999). Modeling time series of count data. In *Asymptotics, nonparametrics, and time series*, volume 158 of *Statistics: Textbooks and Monographs* (pp. 63–113). New York: Dekker. 4

Davis, R. A. & Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika*, 96(3), 735–749. 4

Fokianos, K. (2016). Statistical analysis of count time series models: a GLM perspective. In *Handbook of discrete-valued time series*, Chapman & Hall/CRC Handb. Mod. Stat. Methods (pp. 3–27). CRC Press, Boca Raton, FL. 4

Fokianos, K., Støve, B., Tjøstheim, D., & Doukhan, P. (2020). Multivariate count autoregression. *Bernoulli*, 26(1), 471–499. 4

Gao, X. & Song, P. X. K. (2010). Composite likelihood Bayesian information criteria for model selection in high-dimensional data. *Journal of the American Statistical Association*, 105(492), 1531–1540. 17

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. 4

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. 4

Jung, R. & Tremayne, A. (2011). Useful models for time series of counts or simply wrong ones? *AStA Advances in Statistical Analysis*, 95, 59–91. 4

Karlis, D. (2016). Models for multivariate count time series. In *Handbook of discrete-valued time series*, Chapman & Hall/CRC Handb. Mod. Stat. Methods (pp. 407–424). CRC Press, Boca Raton, FL. 4

Kedem, B. & Fokianos, K. (2002). *Regression models for time series analysis*. Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons. 4

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. 4

Pedeli, X., Davison, A. C., & Fokianos, K. (2015). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *Journal of the American Statistical Association*, 110(511), 1229–1238. 4

Steutel, F. W. & van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 7(5), 893–899. 4

Takeuchi, K. (1976). Distribution of informational statistics and a criterion of model fitting. *Suri Kagaku [Mathematical Sciences] (in Japanese)*, 153, 12–18. 17

Varin, C. & Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3), 519–528. 17

Veraart, A. E. (2019). Modeling, simulation and inference for multivariate time series of counts using trawl processes. *Journal of Multivariate Analysis*, 169, 110–129. 3

…ning operations for modeling time series of counts: a survey. *AStA Advances in Statistical Analysis*, 92, 319–341. 4