Multivariate count time series with flexible autocovariances

Stefanos Kechagias
(joint work with Vladas Pipiras, James Livsey and Robert Lund)

SAS Institute

USF Seminar Series in Analytics
February 24, 2017

Outline

Part 1

- Motivation for a bivariate count time series model
- Review of long memory

Part 2

- Model construction
- Quasi-maximum likelihood inference
- Data application
- Future work

Motivations from data

Examples of correlated counts observed at different time points

- Number of patients with different but related symptoms
- Occurrences of physical phenomena at different locations
- Number of occurrences of various crime types
- Number of trades of different stocks in a portfolio

Capturing both serial and cross-correlation complicates model specification and inference.

Motivations from data

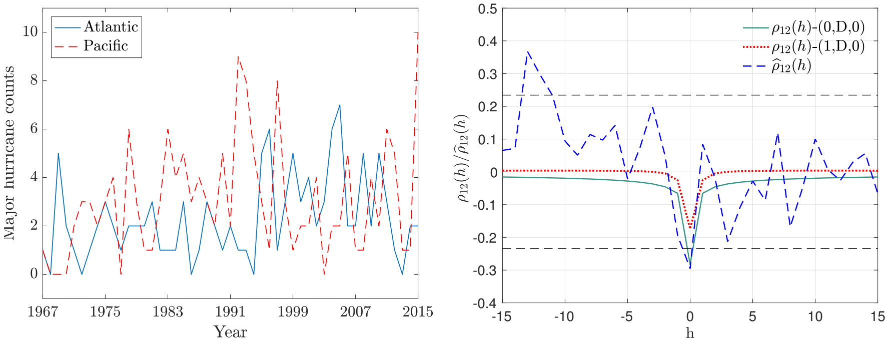


Figure: Left: Annual number of Saffir-Sampson category 3 and stronger hurricanes in North Pacific and North Atlantic Basins. Right: Theoretical/sample (red, green/blue) ccfs of the numbers of major hurricanes in the Atlantic and the Pacific.

- Negative cross correlation
- NA consistent with Poisson assumption, NP somewhat overdispersed

Motivation from data

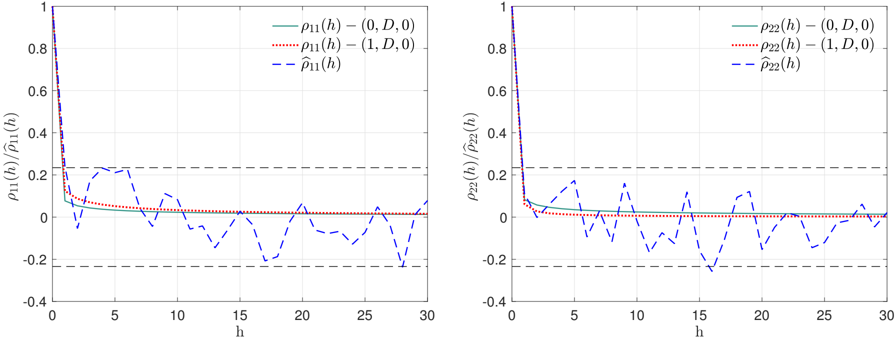


Figure: Sample (blue lines) and theoretical (red, green lines) auto-correlation functions of major hurricane counts in the Atlantic and Pacific Basins.

- Slow decay of dependence?

We would like to have a model that captures all 3 features.

Literature

Bivariate Poisson $BP(\theta_1, \theta_2, \theta_0)$

$$P(y_1, y_2) = e^{-(\theta_1+\theta_2+\theta_0)} \frac{\theta_1{}^{y_1}}{y_1!} \frac{\theta_2{}^{y_2}}{y_2!} \sum_{s=0}^{\min(y_1,y_2)} \binom{y_1}{s} \binom{y_2}{s} s! \left(\frac{\theta_0}{\theta_1 \theta_2}\right)^s$$

- Poisson$(\theta_i + \theta_0)$ marginals, covariance θ_0
- Computationally intensive for large counts, positive correlation
- $BP(\theta_1, \theta_2, 0)$, $(\theta_1, \theta_2) \sim$ logNormal. Positive and negative correlation but marginals require numerical integration
- multivariate case?

Literature

- INAR(1) model

$$Y_t = \alpha \circ Y_{t-1} + R_t,$$

$\alpha \in [0,1]$, R_t uncorrelated, nonnegative, integer-valued random variables. The thinning operator $\circ$ is defined as

$$\alpha \circ Y = \sum_{i=1}^{Y} Z_i, \quad Z_i \sim \text{Bernoulli}(\alpha)$$

- BINAR(1)

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \circ \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} R_{1t} \\ R_{2t} \end{bmatrix}$$

- positive mean, variance and autocovariance

Some notation

- $X = \{X_n\}_{n\in\mathbb{Z}}$ is a zero mean, second-order stationary time series
- $\rho(k) = \mathbb{E}X_nX_{n+k}, k \in \mathbb{Z}$, is the autocorrelation function of X
- $\{\epsilon_n\}_{n\in\mathbb{Z}}$ is a WN series with $\mathbb{E}\epsilon_n^2 = 1$.
- $a_n \sim b_n$ implies that $a_n/b_n \to 1$ as $n \to \infty$.

The definitions and models of the time series we will discuss involve the so-called *long-range dependent* parameter $d \in (0, 1/2)$.

Univariate long-range dependent series

Consider the following two **non-equivalent** conditions:

I. X_n admits the form $X_n = \sum_{k=0}^{\infty} \psi_k \epsilon_{n-k}, \quad \psi_k \underset{k\to\infty}{\sim} c_1 k^{d-1}$

II. The acf of X_n satisfies $\rho(k) \sim c_2 k^{2d-1}$ as $k \to \infty$.

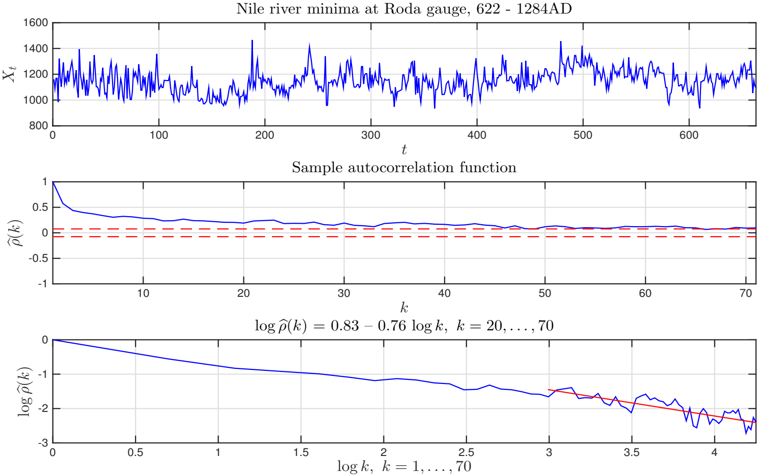
- X_n is called *long-range dependent* if one of the conditions I–II holds.
- Conditions I,II, imply that $\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty$.
- A series with $\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty$ is called *short-range dependent*.

LRD vs SRD

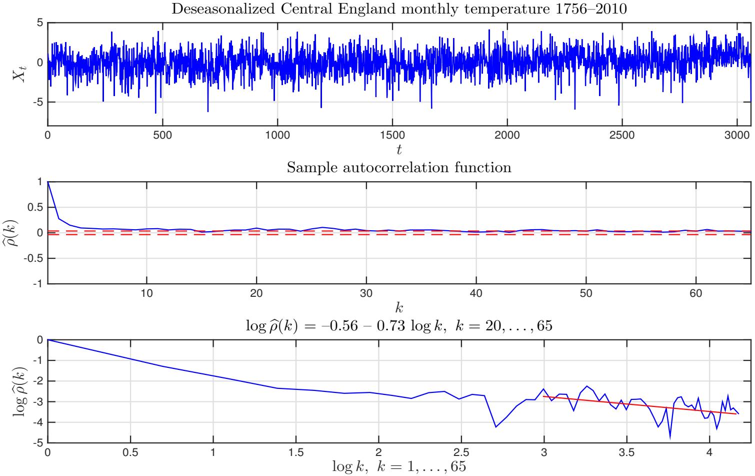
Characterization	LRD	SRD
$\sum_{k=-\infty}^{\infty} \|\rho(k)\|$	∞	$< \infty$
$\rho(k)$	k^{2d-1}	$O(e^{-\alpha k})$
$\sum_{k=0}^{\infty} \psi_k \epsilon_{n-k}$	$\psi_k \sim k^{d-1}$	$O(e^{-\alpha k})$
$\mathrm{Var}(\overline{X}_N)$	N^{2d-1}	N^{-1}

Question: How does the last relation affect inference?

Some examples of LRD series



Some examples of LRD series



ARFIMA$(0, d, 0)$ - A parametric LRD model

- If $d \in \mathbb{Z}^+$, then $\{X_n\}_{n\in\mathbb{Z}}$ is said to be an $ARIMA(0, d, 0)$ process if

$$(I - B)^d X_n = \epsilon_n, \tag{1}$$

- If $d \in (0, 1/2)$ we interpret the solution of (1) as

$$X_n = (I - B)^{-d}\epsilon_n = \sum_{k=0}^{\infty} b_k B^k \epsilon_n = \sum_{k=0}^{\infty} b_k \epsilon_{n-k},$$

- $b_k's$ are the Taylor coefficients of $(1 - z)^{-d}$ and satisfy

$$b_k = \frac{\Gamma(k + d)}{\Gamma(k + 1)\Gamma(d)} \underset{k\to\infty}{\sim} k^{d-1}, \quad \sum_{k=0}^{\infty} b_k^2 < \infty.$$

Bivariate long-range dependent series

- $\{X_n\}_{n\in\mathbb{Z}}=\{(X_{1,n},X_{2,n})'\}_{n\in\mathbb{Z}}$ is a bivariate, zero mean, second-order stationary time series
- $\rho_{ij}(k) = \mathbb{E}X_{i,n}X_{j,n+k},\ i,j = 1,2$ is the (i,j) component of the autocorrelation matrix function
- $\{\epsilon_n\}_{n\in\mathbb{Z}}=\{(\epsilon_{1,n},\epsilon_{2,n})'\}_{n\in\mathbb{Z}}$ is a bivariate WN with $\mathbb{E}\epsilon_n\epsilon_n' = \Sigma$.

The definitions and models of the time series we will discuss involve the so-called *long-range dependent* parameters $d_1, d_2 \in (0, 1/2)$.

Definitions of bivariate LRD series

- Time domain definition of bivariate LRD

$$\rho_{ij}(h) \sim R_{ij} h^{d_i+d_j-1}, \quad h \to \infty,$$

where $R_{ij} \in \mathbb{R}$, $i, j = 1, 2$.

- VARFIMA$(0, D, 0)$ model, $D = \text{diag}(d_1, d_2)$.

$$X_n = \begin{pmatrix} X_{1,n} \\ X_{2,n} \end{pmatrix} = \begin{pmatrix} (I-B)^{-d_1}\epsilon_{1,n} \\ (I-B)^{-d_2}\epsilon_{2,n} \end{pmatrix} = (I-B)^{-D}\epsilon_n,$$

- The acmf has the exact form

$$\rho_{ij}^X(h) = \sigma_{ij} \frac{(-1)^h \Gamma(1-d_i-d_j)}{\Gamma(1-d_i+h)\Gamma(1-d_j-h)}, \; i, j = 1, 2.$$

Bivariate Count LRD model

1. *Let X_t be a bivariate LRD series* with acmf $\rho^X(h), h \in \mathbb{Z}$ and

$$X_t \sim N_2\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \quad \rho = \rho^X(0).$$

2. *Place the components of X into categories*:

$$S_t = (S_{1,t}, S_{2,t})'_{t\in\mathbb{Z}} = \left(1_{\{X_{1,t}>0\}}, 1_{\{X_{2,t}>0\}}\right)'_{t\in\mathbb{Z}},$$

Fact: The series $\{S_t\}_{t\in\mathbb{Z}}$ is stationary with $\mathbb{E}S_t = (1/2, 1/2)'$ and

$$\rho^S_{ij}(h) = \frac{1}{2\pi}\text{asin}(\rho^X_{ij}(h)), \quad h \in \mathbb{Z}.$$

Remark 1: $P(X_{i,t} > 0, X_{j,t+h} > 0) = \frac{1}{4} + \text{asin}(\rho^X_{ij}(h))/2\pi$.

Remark 2: $\text{asin}(z) \underset{z\to 0}{\sim} z$ and so S_t is LRD if X_t is LRD.

Bivariate Count LRD model

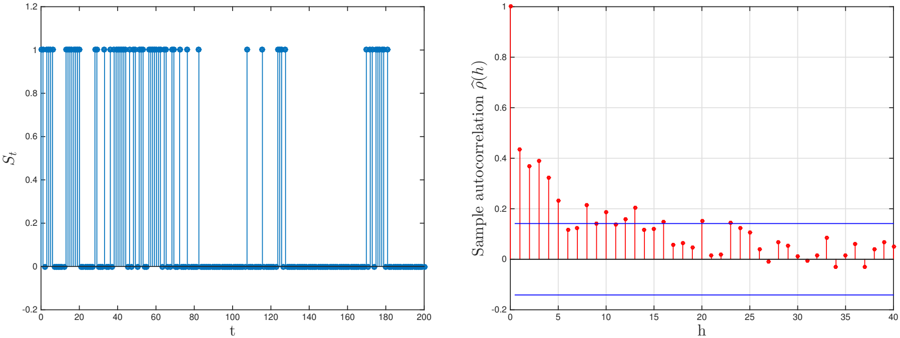


Figure: Left: One realization of the series $S_{1,t}$ with sample size $N = 200$. The underlying process $X_{1,t}$ is a Gaussian FARIMA$(0, 0.4, 0)$. Right: Sample acf $\widehat{\rho}^S(h)$ for lags $h = 0, \dots, 40$.The blue lines indicate the 95% confidence interval.

Binomial marginal distributions

3. *Superimpose IID copies of S_t*

   Binomial marginals

   Let $\{S_{1,t}^{(k)}, S_{2,t}^{(k)}\}_{k=1}^{\infty}$ be a sequence of IID copies of S_t. Then the series Y_t with components

$$Y_{i,t} = \sum_{k=1}^{M} S_{i,t}^{(k)}, \quad i = 1, 2,$$

   has *Binomial*$(M, 1/2)$ univariate marginal distributions with

$$\rho^Y(h) = M\rho^S(h), \quad h \in \mathbb{Z}.$$

   Long memory of S_t implies that Y_t is also LRD.

Poisson marginal distributions

Poisson marginals

Consider the series $\{Y_t\}_{t\in\mathbb{Z}} = \{(Y_{1,t}, Y_{2,t})'\}_{t\in\mathbb{Z}}$ with

$$Y_{i,t} = \sum_{k=1}^{N_{i,t}} S_{i,t}^{(k)}, \quad N_{i,t} \overset{\text{ind}}{\sim} \text{Poisson}(\lambda_i), \quad \lambda_i > 0.$$

Fact: The series Y_t is stationary with $\mathbb{E}Y_t = (\lambda_1/2, \lambda_2/2)'$ and

$$\rho_{ij}^{Y}(h) = \frac{1}{2\pi} c_{ij} \text{asin}(\rho_{ij}^{X}(h)),$$

$$c_{ij} = \begin{cases} 2\lambda_i, & i = j, h = 0, \\ \lambda_i F_{W_{ij}}(-1) + \lambda_j(1 - F_{W_{ij}}(1)), & i \neq j, \end{cases}$$

where $W_{ij} \sim \text{Skellam}(\lambda_i, \lambda_j) = \text{Poisson}(\lambda_i) - \text{Poisson}(\lambda_j)$.

Sketch Proof

Joint pmf of $N_{i,t}, N_{j,t+h}$: $p_{ij}(h) = P(N_{i,t} = n_i, N_{j,t+h} = n_j)$.

$$
\begin{aligned}
\mathbb{E} Y_{i,t} Y_{j,t+h} &= \mathbb{E}\Big(\mathbb{E} \sum_{m=1}^{n_i} S_{i,t}^{(m)} \sum_{k=1}^{n_j} S_{j,t+h}^{(k)} \Big| N_{i,t} = n_i, N_{j,t+h} = n_j\Big) \\
&= \sum_{n_i, n_j = 0}^{\infty} \mathbb{E}\Big(\sum_{m=1}^{n_i} S_{i,t}^{(m)} \sum_{k=1}^{n_j} S_{j,t+h}^{(k)} \Big| N_{i,t} = n_i, N_{j,t+h} = n_j\Big) p_{ij}(h) \\
&= \sum_{n_i, n_j = 0}^{\infty} \Big(M_{ij} \frac{\arcsin(\rho_{ij}^{X}(h))}{2\pi} + \frac{\Pi_{ij}}{4}\Big) p_{ij}(h) \\
&= \tfrac{1}{2\pi} \arcsin(\rho_{ij}^{X}(h)) \mathbb{E} \min(N_{i,t}, N_{j,t+h}) + \tfrac{1}{4} \mathbb{E} N_{i,t} N_{j,t+h}.
\end{aligned}
$$

- $M_{ij} = \min(n_i, n_j)$ cross products of $S_{i,t}^{(m)}$ and $S_{j,t+h}^{(k)}$ for $m = k$
- $\Pi_{ij} - M_{ij}$ cross products of $S_{i,t}^{(m)}$ and $S_{j,t+h}^{(k)}$ for $m \neq k$, $\Pi_{ij} = n_i n_j$

Statistical Inference

Let $\mathbf{Y}_N = (Y_1 \ldots, Y_N)'$ with $\Gamma_N = \mathbb{E}\mathbf{Y}_N\mathbf{Y}_N'$.

- Gaussian log-Likelihood

$$\ell(\Gamma_N) \propto -\frac{1}{2}\log|\Gamma_N| - \frac{1}{2}\mathbf{Y}^T\Gamma_N^{-1}\mathbf{Y}_N$$

- It can be rewritten as

$$\ell(\theta;\mathbf{Y}_N) \propto -\frac{1}{2}\sum_{j=1}^{N}\log|V_{j-1}| - \frac{1}{2}\sum_{j=1}^{N}(Y_j - \widehat{Y}_j)'V_{j-1}^{-1}(Y_j - \widehat{Y}_j),$$

where $\widehat{Y}_j = \mathbb{E}(Y_j|Y_1,\ldots,Y_{j-1})$, $V_{j-1} = \mathbb{E}(Y_j - \widehat{Y}_j)(Y_j - \widehat{Y}_j)'$

- $\widehat{Y}_j$, V_{j-1} can be computed can be computed from $\rho^Y(h)$ in $O(N^2)$.

Statistical Inference

Let $\Phi(z), \Theta(z)$ be the typical AR, MA polynomials with orders p, q

LRD model: $\Phi(B)X_n = (I - B)^{-D}\Theta(B)\epsilon_n$, with $p, q \leq 1$

Parameter: $\theta = (d_1, d_2, \lambda_1, \lambda_2, \rho, \text{vec}(\Phi), \text{vec}(\Theta))$.

Quasi-MLE: $\widehat{\theta} = \underset{\theta}{\operatorname{argmax}}\ \ell(\theta; Y)$

Computational issues

- The entries of Σ, Φ, Θ appear in nonlinear constraints.
- We assumed marginal unit variances and a prescribed correlation for X. What parameters σ_{ij} ensure that?
- What if the desired σ_{ij} do not satisfy the constraints?

Simulation-$X_t \sim$VARFIMA$(0, D, 0)$

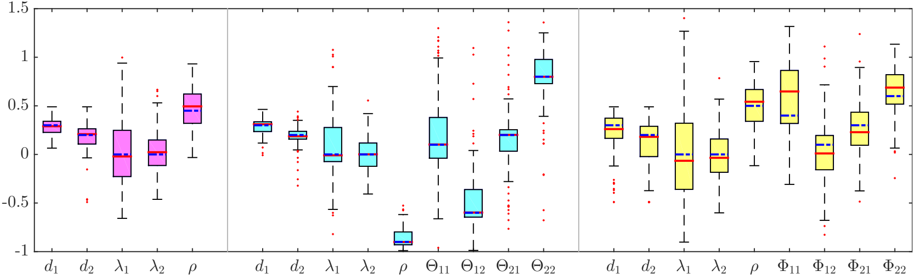


Figure: Boxplots of the estimates for $(p, q) = (0, 0)_1$ (left box), $(0, 1)$ (middle box) and $(1, 0)$ (right box) $T = 400$ and 100 replications. The dashed blue lines correspond to the true parameter values, while the solid red lines are the medians.

Simulation results

(p,q)	$(0,0)$		$(1,0)$		$(0,1)$	
T	200	400	200	400	200	400
d_1	-0.011 0.082	-0.010 0.059	-0.029 0.138	-0.039 0.107	0.059 0.057	0.069 0.038
d_2	-0.026 0.093	0.009 0.065	-0.019 0.183	-0.190 0.179	0.071 0.070	0.085 0.047
λ_1	-0.046 0.273	-0.021 0.229	-0.049 0.309	-0.065 0.338	-0.024 0.231	-0.013 0.199
λ_2	0.029 0.157	0.023 0.125	-0.077 0.228	-0.035 0.165	-0.023 0.156	0.010 0.149
ρ	0.072 0.218	0.045 0.155	-0.002 0.196	0.042 0.163	0.060 0.110	-0.011 0.050
$\Phi_{1,1}/\Theta_{1,1}$			0.069 0.243	0.248 0.267	-0.004 0.021	-0.017 0.023
$\Phi_{1,2}/\Theta_{1,2}$			0.017 0.241	-0.090 0.178	0.165 0.085	0.145 0.086
$\Phi_{2,1}/\Theta_{2,1}$			-0.023 0.158	-0.071 0.168	0.130 0.069	0.133 0.047
$\Phi_{2,2}/\Theta_{2,2}$			0.052 0.148	0.088 0.140	0.029 0.020	0.032 0.022

Table: MB and MAD of estimated parameters. True parameter values are $d_1 = 0.3, d_2 = 0.2, \lambda_1 = 3, \lambda_2 = 2\ \rho = -0.9,\ \Phi_{1,1} = 0.4, \Phi_{1,2} = 0.1,\ \Phi_{2,1} = 0.3,$ $\Phi_{2,2} = 0.6,\ \Theta_{1,1} = 0.1, \Theta_{1,2} = -0.6,\ \Theta_{2,1} = 0.2, \Theta_{2,2} = 0.8.$

Data application

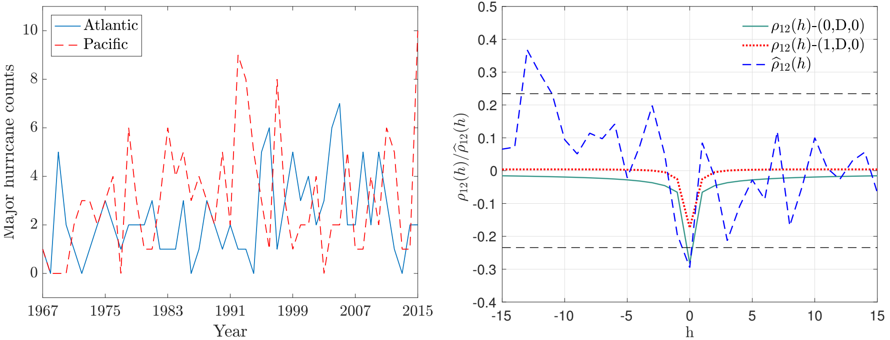


Figure: Left: Annual number of Saffir-Sampson category 3 and stronger hurricanes in NP and NA Basins. Right: Theoretical/sample ccfs

- $\bar{x}_{\text{Atl}} = 2.31, s^2_{\text{Atlantic}} = 2.97,\ \bar{x}_{\text{Pac}} = 3.10, s^2_{\text{Pacific}} = 5.76$
- $\text{IC}_{(0,D,0)} < \text{IC}_{(1,D,0)}$. For $(1, d, 0)$ $|\Sigma| < 0$ when $\rho < -0.7$.
- $\widehat{d}_1 = 0.24, \widehat{d}_2 = 0.23, \widehat{\lambda}_1 = 5.7, \widehat{\lambda}_2 = 11.5, \widehat{\rho} = -0.96$, $(\widehat{\rho}_Y = -0.28)$

Other marginal distributions

1. If $N_{i,t}$ ~Geo(α) then $Y_{i,t}$ ~Geo(γ), $\alpha = 2\gamma/(\gamma + 1)$.

2. Using the fact above we can also obtain a *NB* marginal.

3. For DU($\{u_1, \ldots, u_D\}$) pick $r_1, \ldots, r_D$

$$P(r_{k-1,i} < X_{i,t} < r_{k,i}) = \frac{1}{D}, \tag{2}$$

and take the series

$$U_{t,i} = \sum_{k=0}^{D-1} u_k 1_{[r_{k,i} < X_{i,t}. < r_{k+1,i}]}$$

For the latter we cannot use the quadrant Normal probabilities to obtain an exact form for the acv function.

Conclusions - Future work

1. Negative correlation, arbitrary marginals, LRD behavior
2. From bivariate to multivariate?
3. Other estimation methods (Pseudo-Likelihood, Bayesian, etc.)
4. Performance under misspecification and forecasting power

References

- Beran, et al., *Long-Memory Processes*. Springer, Heidelberg, 2013.
- Davis et al., eds. *Handbook of Discrete-Valued Time Series*. CRC Press, 2016.
- Livsey et al., *Multivariate count time series with flexible autocovariances*. Submitted