SAND2018-6765C

Temporal latent processes for heavy-tailed data

PRESENTED BY

Gabriel Huerta[1]

[1] Department of Statistical Sciences
Sandia National Laboratories

*co-author: Kellin Rumsey. (UNM/SNL)

ries is a multimission laboratory managed and operated by National Technology & Engineering Solutions of San
well International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under cont
tional Nuclear Security Administration
under contract DE-NA-0003525. SAND NO.

Outline

- Background
- Our modeling approach.
- Comparison to other processes
- Numerical illustrations.
- Extensions.

General setup

- Assume we have a sequence of measurements X_t.
- Fix a threshold u and model values above threshold,

$$Y_t = \max(X_t - u, 0)$$

and

$$Z_t = \begin{cases} 1, & \text{if} \quad X_t > u \\ 0, & \text{otherwise} \end{cases}$$

- Classical EVT (Pickands 1975) justify modeling $Y_t | Z_t = 1$ with a **Generalized Pareto Distribution,**

$$f(y|\xi, \sigma) = \frac{1}{\sigma}\left(1 + \xi \frac{y}{\sigma}\right)^{-(1/\xi+1)} ; y \geq 0$$

- ξ is the *shape* parameter; σ is the *scale* parameter.

Previous work

- Most approaches treat X_t as i.i.d and consider GPD.
- Coles (2001); Coles and Powell (1996) overview Bayesian analysis and classical estimation.
- Account for temporal and/or spatial dependence.
    - Majorly done with assuming parameters follow a temporal/spatial process.
    - Casson and Coles, (1999); Gaetan and Grigoletto,(2004); Huerta and Sanso, (2007); Sang and Gelfand, (2009) and many more
    - Copula based: Nakijama (2001,2015), Ning and Bloomfield (2017).
    - *Latent process*:. Gaetan and Bortot (2014); Nieto-Barajas and Huerta (2017)

GPD as a scale mixture

- Consider a re-parameterized model with $\alpha = 1/\xi$ and $\beta = \sigma/\xi$
- The GPD (α, β) has the form

$$f(y|\alpha,\beta) = \frac{\alpha}{\beta}\left(1+\frac{y}{\beta}\right)^{-(\alpha+1)} \quad ; y \geq 0$$

- For $\alpha \geq 0$ following Ries and Thomas (2007),

$$f(y|\alpha,\beta) = \int_0^\infty \lambda e^{-\lambda y} g(\lambda|\alpha,\beta) d\lambda$$

- where

$$g(\lambda|\alpha,\beta) = [\beta^\alpha/\Gamma(\alpha)]\lambda^{\alpha-1}e^{-\beta\lambda}$$

- $\alpha \geq 0$ restricts the *Domain of attraction* to Fréchet.

The modeling approach (temporal case)

- We have sequence of observations Y_t where

$$P(Y_t > 0|\Lambda_t) = e^{-\kappa\Lambda_t}$$

and $Y_t = 0$ otherwise.

- For $Y_t > 0$, we use a mixture of Exponential and Gamma variables.

$$\begin{aligned} Y_t|\Lambda_t &\sim Exp(\Lambda_t) \\ \Lambda_t|\mathbf{Z} &\sim Ga\left(\alpha + \sum_{s=0}^{p} Z_{t-s}, \beta + \sum_{s=0}^{p} \gamma_{t-s}\right). \end{aligned}$$

- To model the latent variables $\mathbf{Z}$ and parent W

$$\begin{aligned} Z_t|W &\sim Po(\gamma_t W) \\ W &\sim Ga(\alpha, \beta). \end{aligned}$$

- Marginally Y_t follows a GPD with $\alpha = 1/\xi$ and $\beta = \sigma/\xi$.

Customized MCMC approach

- Joint posterior

$$\prod_{t=1}^{T} f(Y_t|\Lambda_t) \prod_{t=1}^{T} f(\Lambda_t|\mathbf{Z}) \prod_{t=1}^{T} f(Z_t|W) f(W|\alpha, \beta) p(\alpha) p(\beta) p(\gamma)$$

- We define, $\zeta_t = \sum_{s=0}^{p} Z_{t-s}$, $\Gamma_t = \sum_{s=0}^{p} \gamma_{t-s}$ and $p_t = e^{-\kappa \Lambda_t}$.
- We iterative sample,

$$\begin{aligned}
p(W|rest) &= Ga\left(\alpha + \sum_{t=1}^{T} Z_t, b + \sum_{t=1}^{T} \gamma_t\right) \\
p(\alpha|rest) &\propto \frac{\left(W\beta \prod_{t=1}^{T} (\beta + \Gamma_t)\right)^{\alpha}}{\Gamma(\alpha) \prod_{t=1}^{T} \Gamma(\alpha + \zeta_t)} p(\alpha) \\
p(\beta|rest) &\propto \prod_{t=1}^{T} \left\{ (\beta + \Gamma_t)^{\alpha + \zeta_t} \right\} e^{-\beta(W + \sum_{t=1}^{T} \Lambda_t)} \beta^{\alpha} p(\beta)
\end{aligned}$$

MCMC approach

- Sample from,

$$p(Z_t|rest) \propto \prod_{s=0}^{p} \left\{ \frac{(\Lambda_{t+s}(\beta+\Gamma_{t+s}))^{\zeta_{t+s}}}{\Gamma(\alpha+\zeta_{t+s})} \right\} \frac{(\gamma_t W)^{Z_t}}{Z_t!}$$

$$p(\gamma_t|rest) \propto \prod_{s=0}^{p} \left\{(\beta+\Gamma_{t+s})^{\alpha+\zeta_{t+s}}\right\} \exp(-(W\gamma_t + \sum_{s=0}^{p} \Gamma_{t+s}\Lambda_{t+s}))\gamma_t^{Z_t}$$

- For $t = 1, 2 \cdots T$, we simulate Λ_t by
  a If $Y_t > 0$, then

$$\Lambda_t|rest \sim Ga(\alpha + \zeta_t + 1, \beta + \Gamma_t + Y_t + \kappa)$$

  b If $Y_t = 0$, then we simulate Λ_t by inverting it's CDF.

Sample estimator for κ

- Recall that $p_t = P(Y_t > 0) = e^{-\kappa\Lambda_t}$
- Thus, we simply use

$$\hat{p}_t = \frac{\sum_{t=1}^{T} I(Y_t > 0)}{T}$$

- Since $\Lambda_t \sim Ga(\alpha, \beta)$, we simply solve

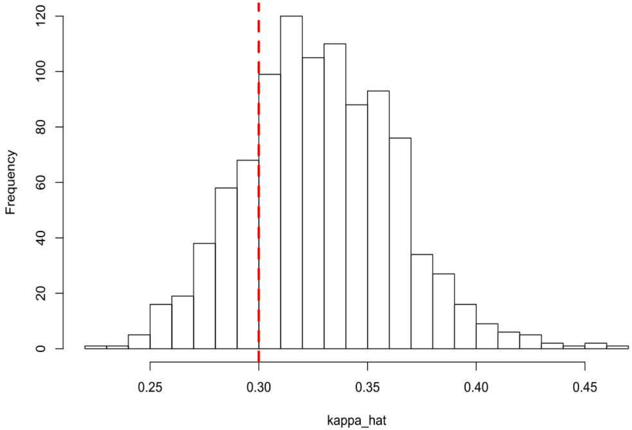
$$\kappa = (\alpha/\beta)^{-1} \log(1/\hat{p}_t)$$

- Since for $Y_t | Y_t > 0$, are marginally $GPD(\xi = 1/\alpha, \sigma = \beta/\alpha)$,

$$\hat{\kappa} = \hat{\sigma} \log(1/\hat{p}_t)$$

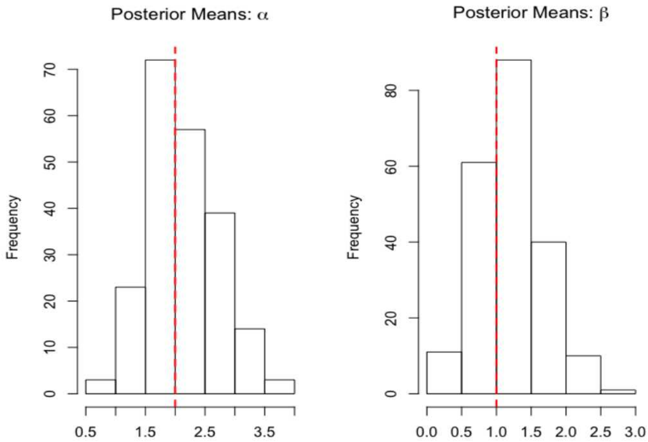
- Studying ways to estimate κ in *good* way.
- Dealing with *threshold estimation* is not straightforward, Scarrott and MacDonald (2012).

Simulation distribution for $\hat{\kappa}$



- True value dashed red line.
- As true κ increases there is a large sample variability.

Simulation study



- Based on 200 experiments, $T = 100$,
- 10000 MCMC iterations.

Gaetan and Bortot (2014) models

- Warren process:

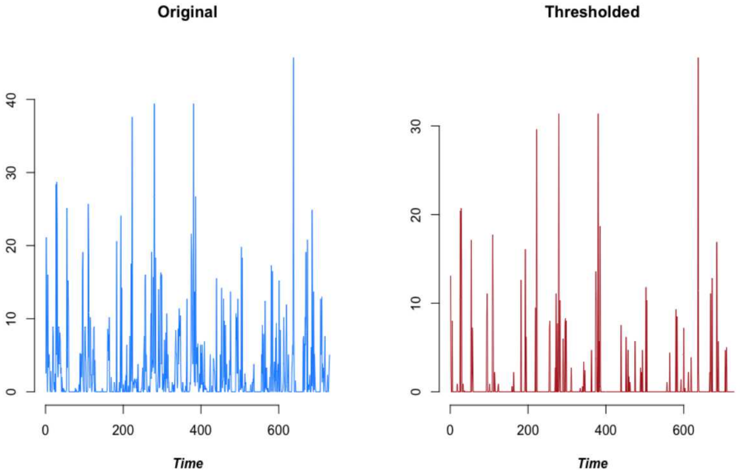
$$\begin{aligned} \Lambda_t | \Pi_t &\sim Ga(\Pi_t + \alpha, (1-\rho)/\beta) \\ \Pi_t | \Lambda_{t-1} &\sim Po\left(\frac{\rho \Lambda_{t-1}}{(1-\rho)}\right); 0 \leq \rho \leq 1 \\ \Lambda_{t-1} &\sim Ga(\alpha, \beta). \end{aligned}$$

- Gaver and Lewis process:

$$\begin{aligned} \Lambda_t &= \rho \Lambda_{t-1} + W_t \\ W_t | \Pi_t &\sim Ga(\Pi_t, \beta/\rho) \\ \Pi_t | \Lambda_{t-1} &\sim Po\left(\frac{\rho \Lambda_{t-1}}{(1-\rho)}\right); 0 \leq \rho \leq 1 \\ P_t &\sim Ga(\alpha, 1) \\ \Lambda_{t-1} &\sim Ga(\alpha, \beta). \end{aligned}$$

- In G&B 14, both are fitted with *pairwise likelihood inference*.

Rain data set from Coles (2001)



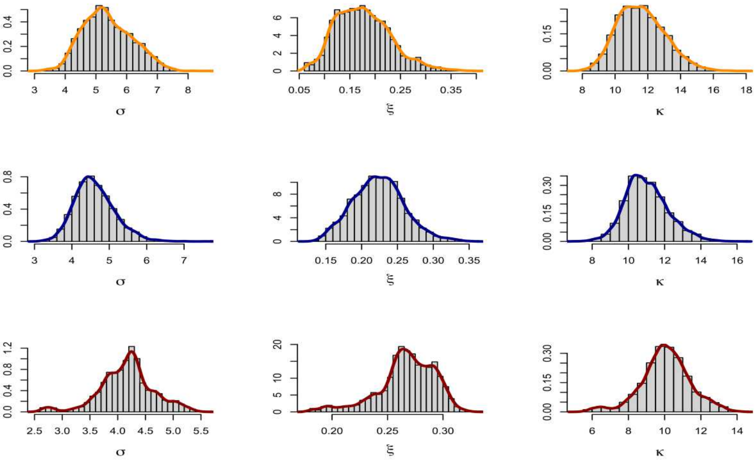
- Daily rainfall in southwest England (2 years, left).
- Suggested threshold, $u = 8$ so $Y_t = X_t - 8$ (right).

Comparison of estimates for 3 methods

Parameter	α	β	W or ρ	κ
Latent-AR process	6.237	34.646	0.189	11.604
	(2.083)	(15.909)	(0.085)	(1.445)
Warren process	4.586	21.196	0.863	11.010
	(0.759)	(4.608)	(0.055)	(1.190)
Gaver-Lewis process	3.779	15.59	0.723	10.069
	(0.407)	(1.281)	(0.002)	(1.362)

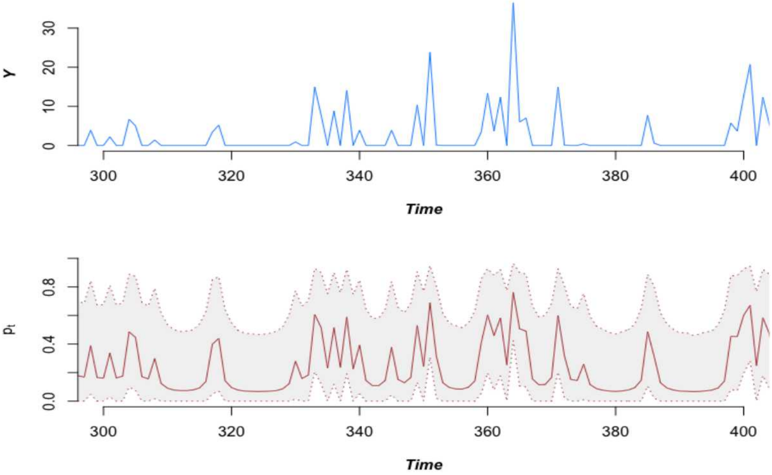
- Posterior mean estimates and SDs in ().
- ρ parameter for G-L process favors lower values.

Posterior distributions for σ, ξ and κ



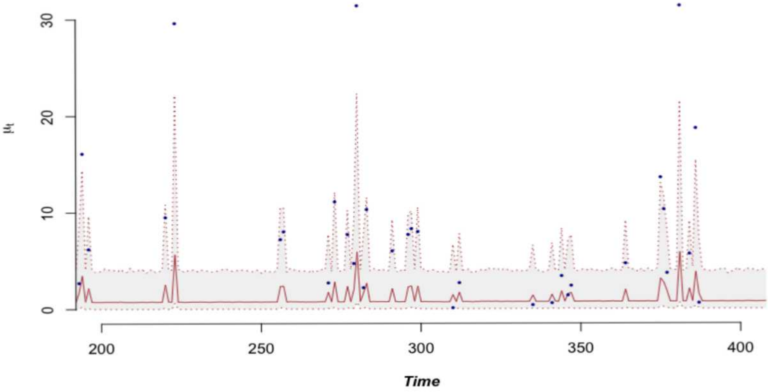
- 1st row: Latent-AR; 2nd: Warren; 3rd: Gaver-Lewis.

Posterior distribution of p_t under latent-AR.



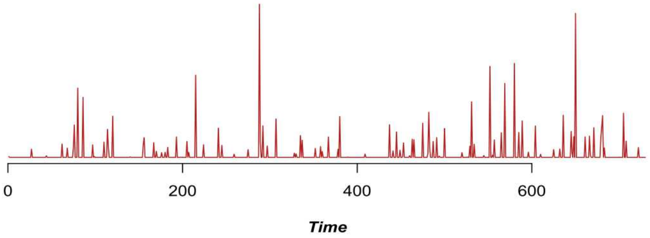
- Solid line: posterior mean of p_t.
- Grey denotes 95% intervals.

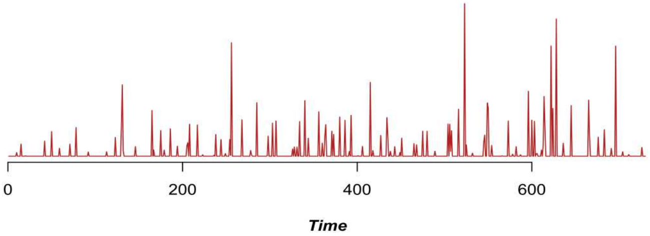
Posterior distribution of $\mu_t = E(Y_t|data)$



- Solid line: posterior mean of μ_t.
- Grey denotes 95% intervals.

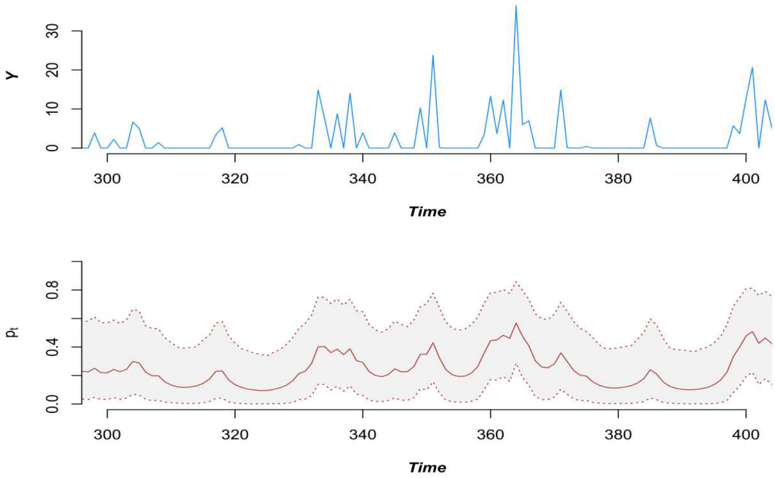
Realizations of Latent-AR process





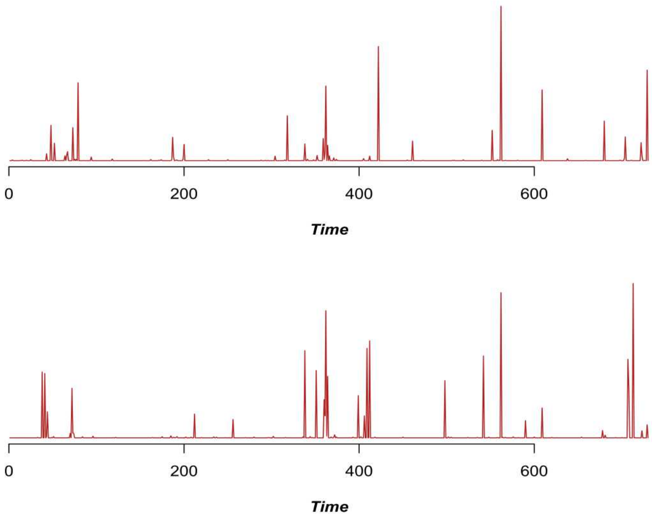
- Based on samples of λ_t and p_t

Posterior distribution of p_t under W-process



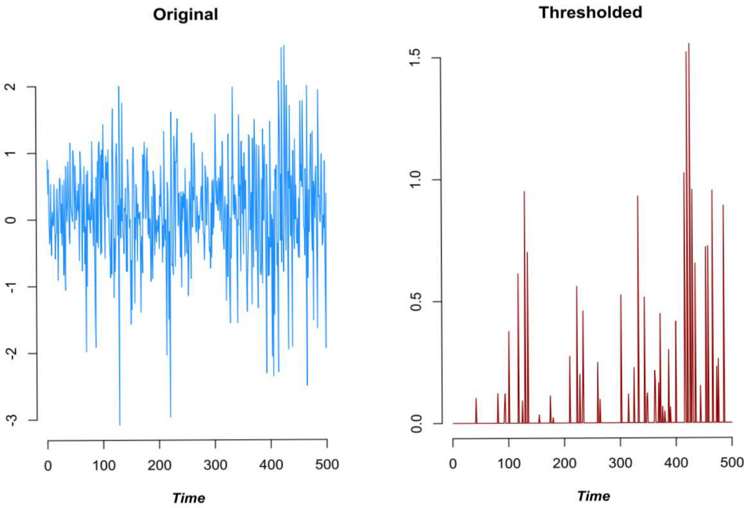
- Solid line: posterior mean of p_t.
- Grey denotes 95% intervals.

Realizations of Warren process



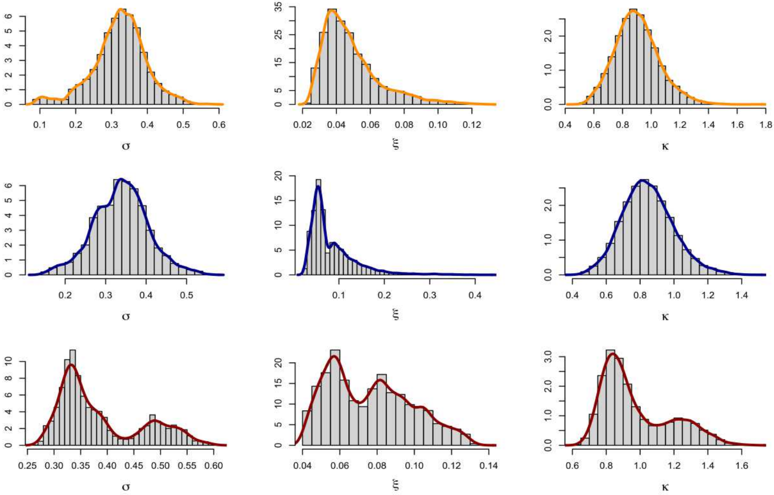
- Based on samples of λ_t and p_t

Daily Closing Prices of the Dow Jones



- Original are *return* values, $X_t = 100 * log(D_t/D_{t-1})$.
- Values thresholded with 0.9 quantile.

Posterior distributions for σ, ξ and κ



- 1st row: Latent-AR; 2nd: Warren; 3rd: Gaver-Lewis.

Comparison of estimates for 3 methods

Parameter	α	β	W or ρ	κ
Latent-AR process	22.751	7.402	3.068	0.903
	(6.681)	(2.77)	(1.341)	(0.154)
Warren process	15.342	5.421	0.021	0.821
	(7.046)	(2.947)	(0.02)	(0.127)
Gaver-Lewis process	14.237	5.495	0.041	0.974
	(4.135)	(1.763)	(0.01)	(0.203)

- Posterior mean estimates and SDs in ().

Some properties

- For the *Latent-AR* model

$$Corr(\Lambda_t, \Lambda_{t+j}) =$$

$$\alpha \left\{ \frac{(\beta^2 + \beta(\Gamma_t + \Gamma_{t+j}) + \mu_1\mu_3) + \mu_2 \left(\mu_1 + \mu_2 + \mu_3 + \frac{\beta+\mu_2}{\alpha}\right)}{(\beta + \Gamma_t)(\beta + \Gamma_{t+j})} - 1 \right\}$$

where $\Gamma_t = \sum_{s=0}^{p} \gamma_{t-s}$ and μ_i are also functions of γ_t.

- For both the W and G-P process,

$$Corr(\Lambda_t, \Lambda_{t+j}) = \rho^j$$

- Since models have $Ga(\alpha, \beta)$ margins,

$$LP^{(1)}(x) = E(e^{-x\Lambda_t}) = \left(\frac{\beta}{\beta + x}\right)^{\alpha}$$

- if $x = u$ we get the probability of being above threshold.

Final Comments

- Bayesian inference of hierarchical models for heavy-tail data.
- Marginally models follow a *GPD*.
- Models applicable to real data.
- Work in progress
  - Assess efficiency for G-L process
  - Extend to spatial or spatial temporal cases.

$$\Lambda_t \rightarrow \Lambda_{t,x}$$

  - Areal case. May follow Nieto-Barajas/Huerta 2017
  - Point reference: Bacron et al. (2017).