Heavy-Tailed Time Series: Theory and Applications

Richard A. Davis
Columbia University
www.stat.columbia.edu/~rdavis

Slides at

http://www.stat.columbia.edu/~rdavis/lectures/Bogota2014_L1-3.pdf

Outline

- ☞ Financial time series modeling
  - General comments
  - Characteristics of financial time series
  - Examples (exchange rate, Amazon)
  - Multiplicative models for log-returns (GARCH, SV)
- ☞ Regular variation
  - univariate case
  - multivariate case
- ☞ Applications of regular variation
  - Stochastic recurrence equations (GARCH)
  - Stochastic volatility
  - Point process convergence
  - Extremes and extremal index
  - Limit behavior of sample correlations

Outline

☞ The Extremogram

- Examples
- Sufficient conditions for existence: regular variation
- Illustrations (permutation procedures)
- Connections with Return Times of Rare Events
- Cross-extremogram (devolatilizing/deGARCHing)
- Bootstrapping the Extremogram
- Theory & examples

Financial Time Series Modeling

One possible goal: Develop models that capture essential features of financial data.

Strategy: Formulate families of models that at least exhibit these key characteristics. (e.g., GARCH and SV)

Linkage with goal: Do fitted models actually capture the desired characteristics of the real data?

Answer wrt to GARCH and SV models: Yes and no. Answer may depend on the features.

Stărică's paper: "Is GARCH(1,1) Model as Good a Model as the Nobel Accolades Would Imply?"

This paper discusses inadequacy of GARCH(1,1) model as a "data generating process" for the data.

Extremes and Time Series Modeling

Two strategies for thinking about modeling extremes in time series:

1. Fit a model to the entire data set (e.g., GARCH and SV for financial time series) and study the extreme value behavior associated with the fitted model as truth.
2. Construct and fit models only to the ***extremes*** (e.g., observations exceeding a large threshold).

Do fitted models actually capture the desired (*extremal)* characteristics of the data?

- How do we assess "fitted" (expected) with "observed"?
- Need a mechanism for measuring extremal dependence.

The ***extremogram*** was developed with these ideas in mind. Ultimately, we might judge goodness of fit by examining how well the "sample" extremogram matches up with the "population" extremogram.

Characteristics of financial time series

Define $X_t = \ln(P_t) - \ln(P_{t-1})$ (log returns)

- heavy tailed

$$P(|X_1| > x) \sim RV(-\alpha), \quad 0 < \alpha < 4.$$

- uncorrelated

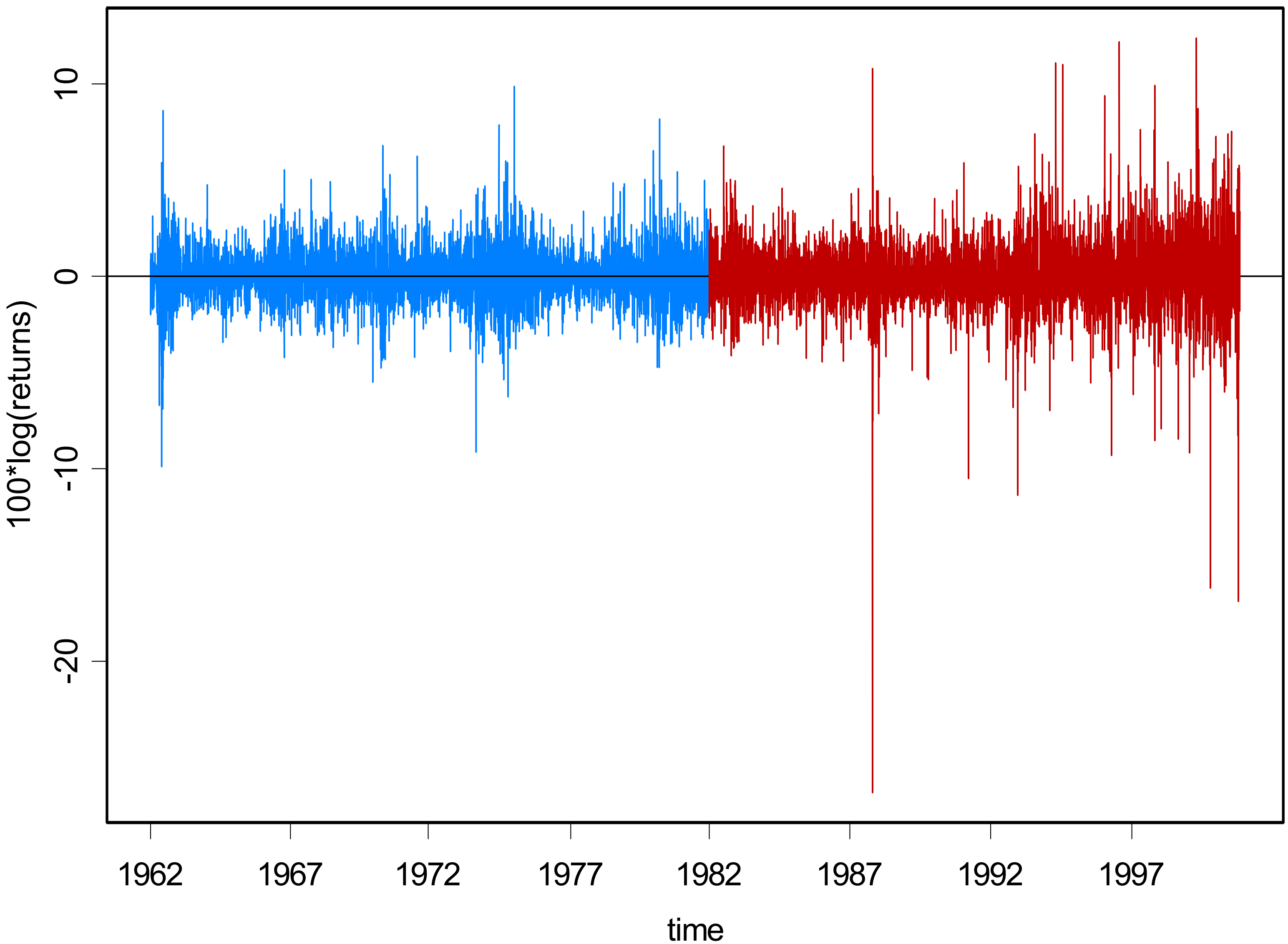
$\hat{\rho}_X(h)$ near 0 for all lags $h > 0$

- $|X_t|$ and X_t^2 have slowly decaying autocorrelations

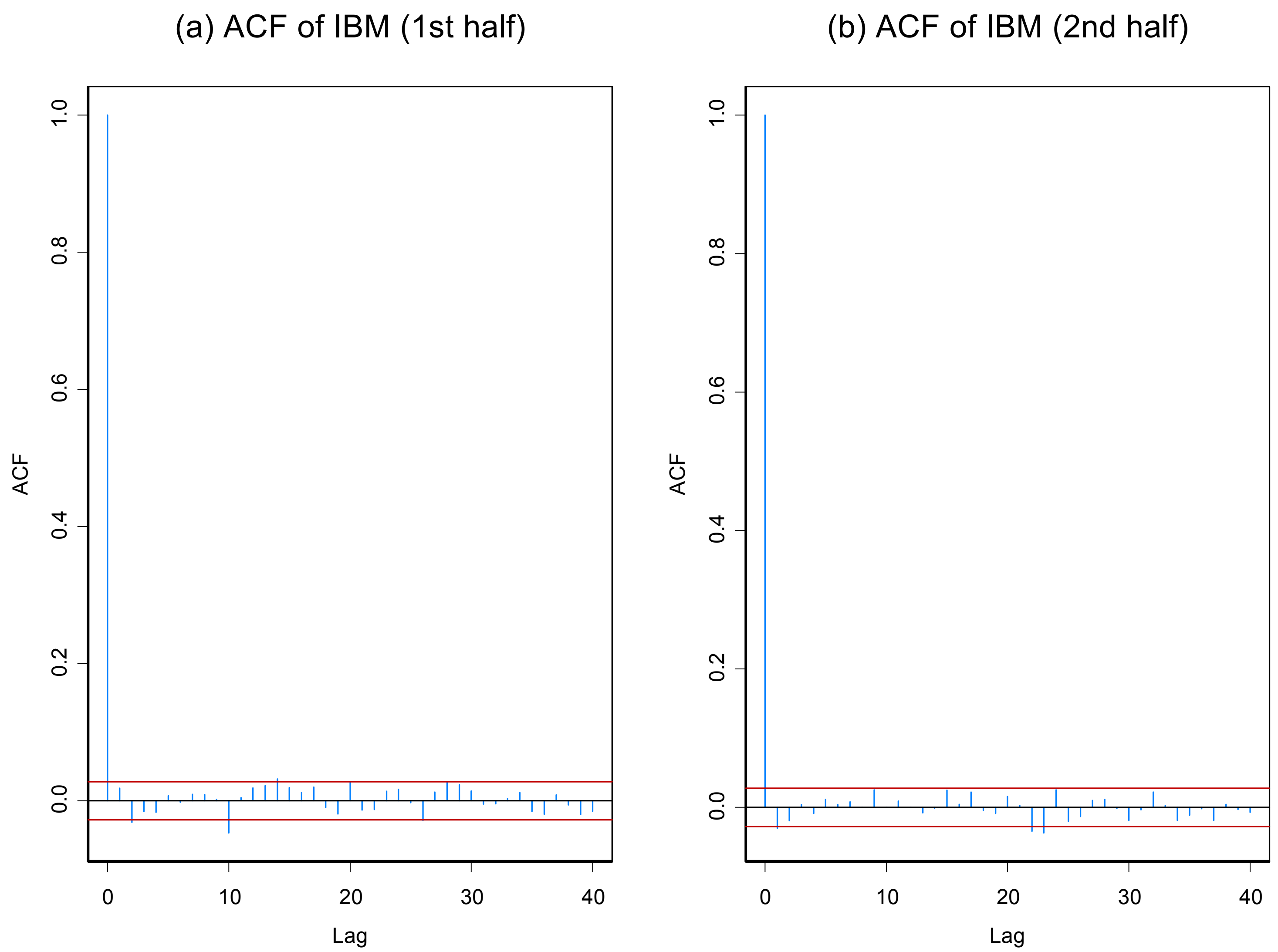
$\hat{\rho}_{|X|}(h)$ and $\hat{\rho}_{X^2}(h)$ converge to 0 *slowly* as h increases.

- process exhibits ‘volatility clustering’.

Log returns for IBM 1/3/62-11/3/00 (blue=1961-1981)

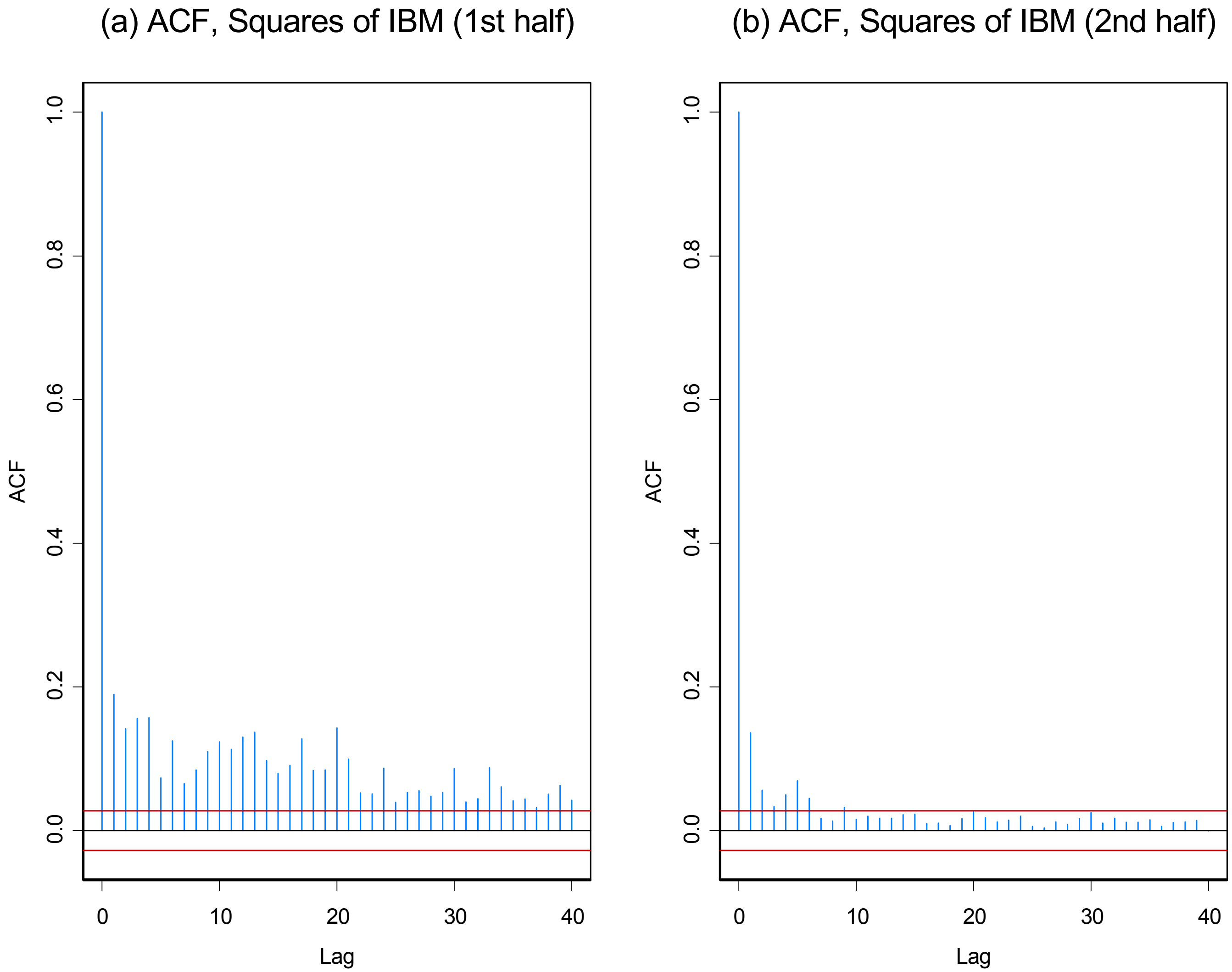


Sample ACF IBM (a) 1962-1981, (b) 1982-2000



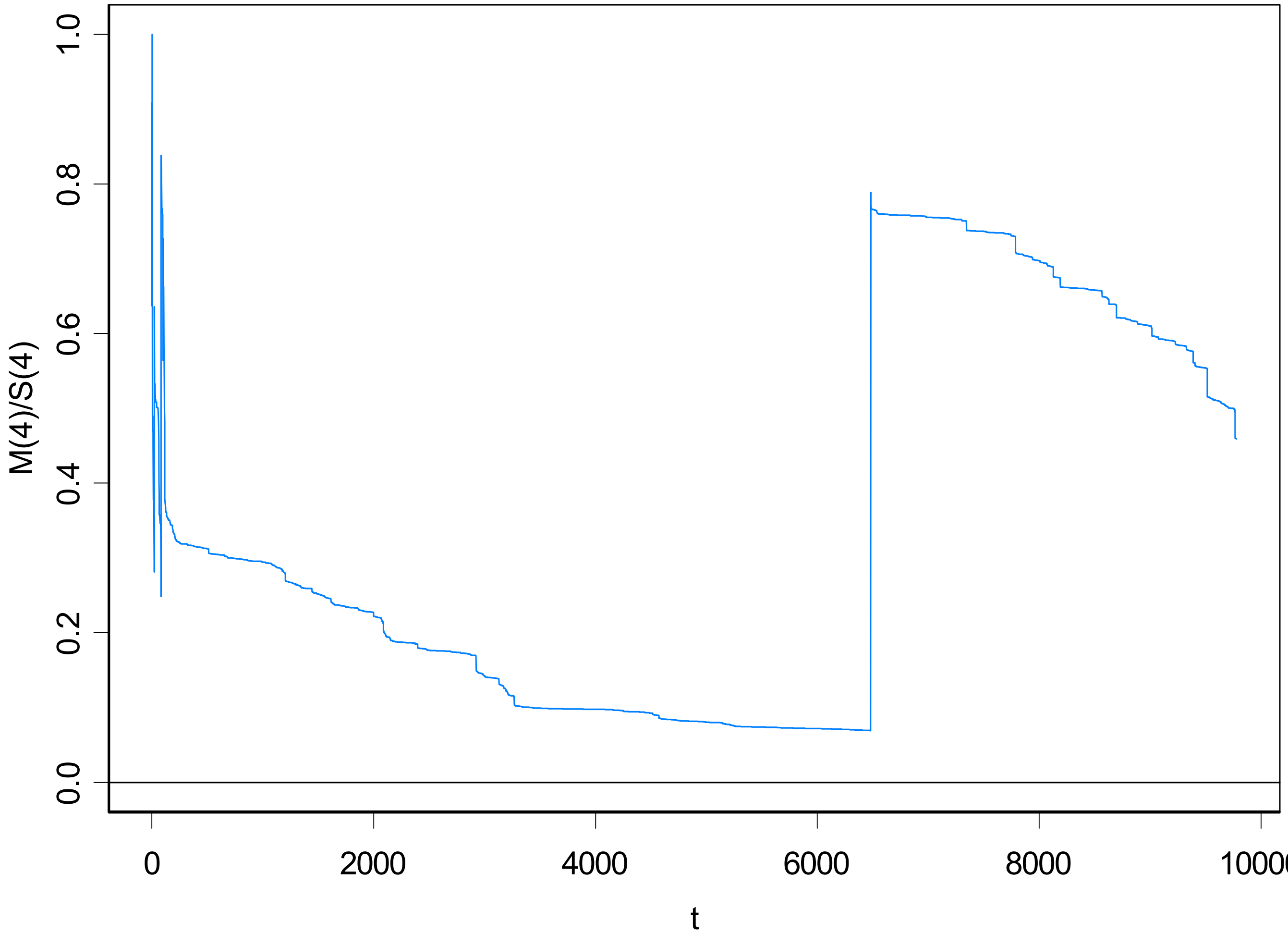
Remark: Both halves look like white noise.

ACF of squares for IBM (a) 1961-1981, (b) 1982-2000



Remark: Series are not independent white noise?

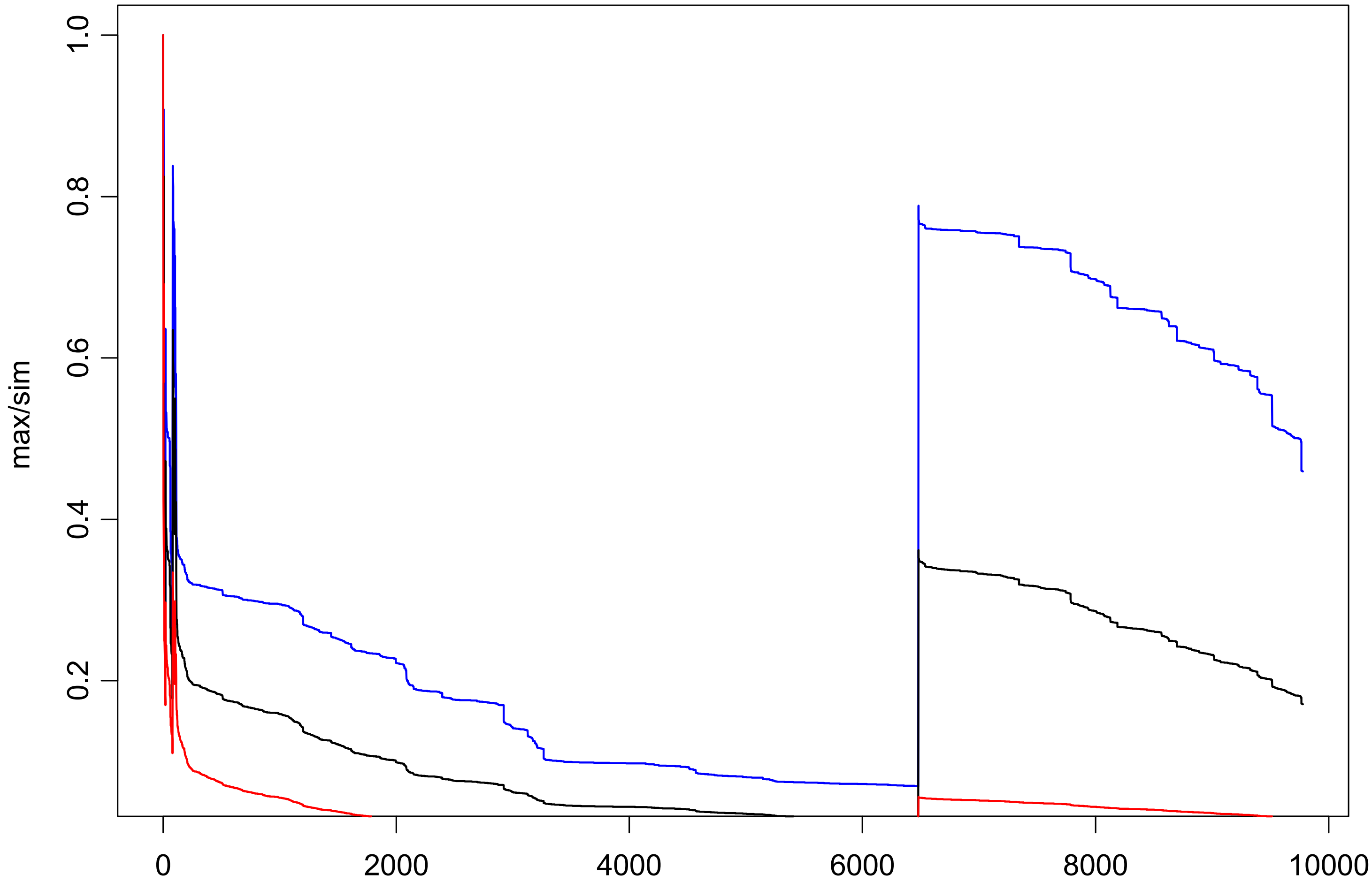
Plot of $M_t(4)/S_t(4)$ for IBM



Remark: For IID data, $M_t(k)/S_t(k) \to 0$ as $t \to \infty$ iff $E|X_1|^k < \infty$, where

$$M_t = \max_{s=1,\ldots,t} |X_s|^k \quad \text{and} \quad S_t = \sum_{s=1}^{t} |X_s|^k$$

Plot of $M_t(k)/S_t(k)$ for IBM, k=4, 3, 2



Remark: For IID data, $M_t(k)/S_t(k) \to 0$ as $t \to \infty$ iff $E|X_1|^k < \infty$, where

$$M_t = \max_{s=1,\ldots,t} |X_s|^k \quad \text{and} \quad S_t = \sum_{s=1}^{t} |X_s|^k$$

Hill's estimator of tail index

The marginal distribution *F* for heavy-tailed data is often modeled using *Pareto-like tails*,

$$1-F(x) = x^{-\alpha}L(x),$$

for *x* large, where $L(x)$ is a slowly varying function ($L(xt)/L(x) \to 1$, as $x \to \infty$). Now if

$$X \sim F, \text{ then } P(\log X > x) = P(X > \exp(x)) = \exp(-\alpha x)L(\exp(x)),$$

and hence log X has an approximate exponential distribution for large *x*. The spacings,

$$\log(X_{(j)}) - \log(X_{(j+1)}),\ j=1,2,\ldots,m,$$

from a sample of size *n* from an exponential distr are approximately independent and $Exp(\alpha j)$ distributed. This suggests estimating α^{-1} by

$$\hat{\alpha}^{-1} = \frac{1}{m}\sum_{j=1}^{m}\left(\log X_{(j)} - \log X_{(j+1)}\right)j$$

$$= \frac{1}{m}\sum_{j=1}^{m}\left(\log X_{(j)} - \log X_{(m+1)}\right)$$

Hill's estimator of tail index

Def: The *Hill estimate* of α for heavy-tailed data with distribution given by

$$1-F(x) = x^{-\alpha}L(x),$$

is

$$\hat{\alpha}^{-1} = \frac{1}{m}\sum_{j=1}^{m}\left(\log X_{(j)} - \log X_{(j+1)}\right)j$$

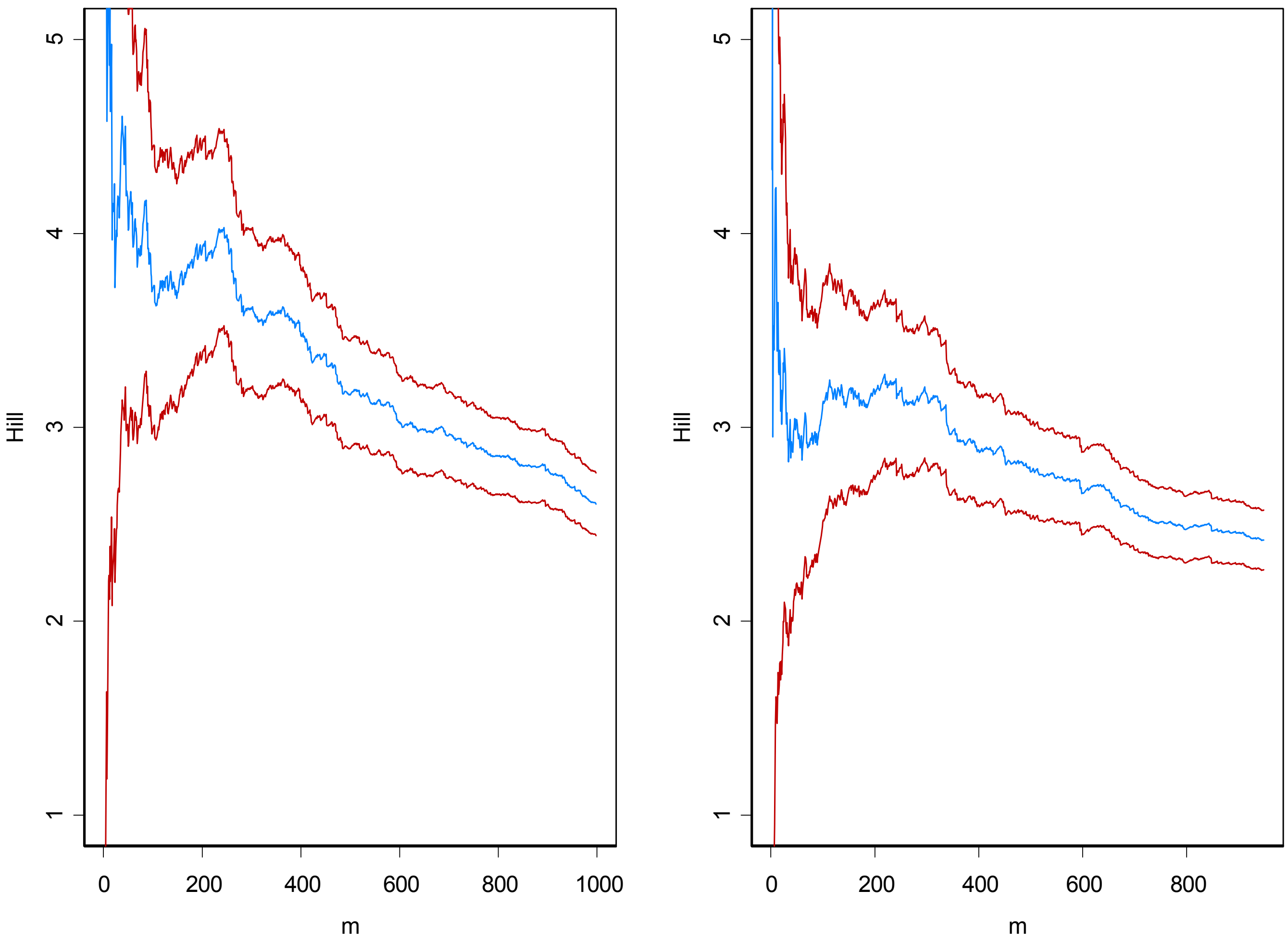
$$= \frac{1}{m}\sum_{j=1}^{m}\left(\log X_{(j)} - \log X_{(m+1)}\right)$$

The *asymptotic variance* of this estimate for α is

α^2/m and estimated by $\hat{\alpha}^2/m$.

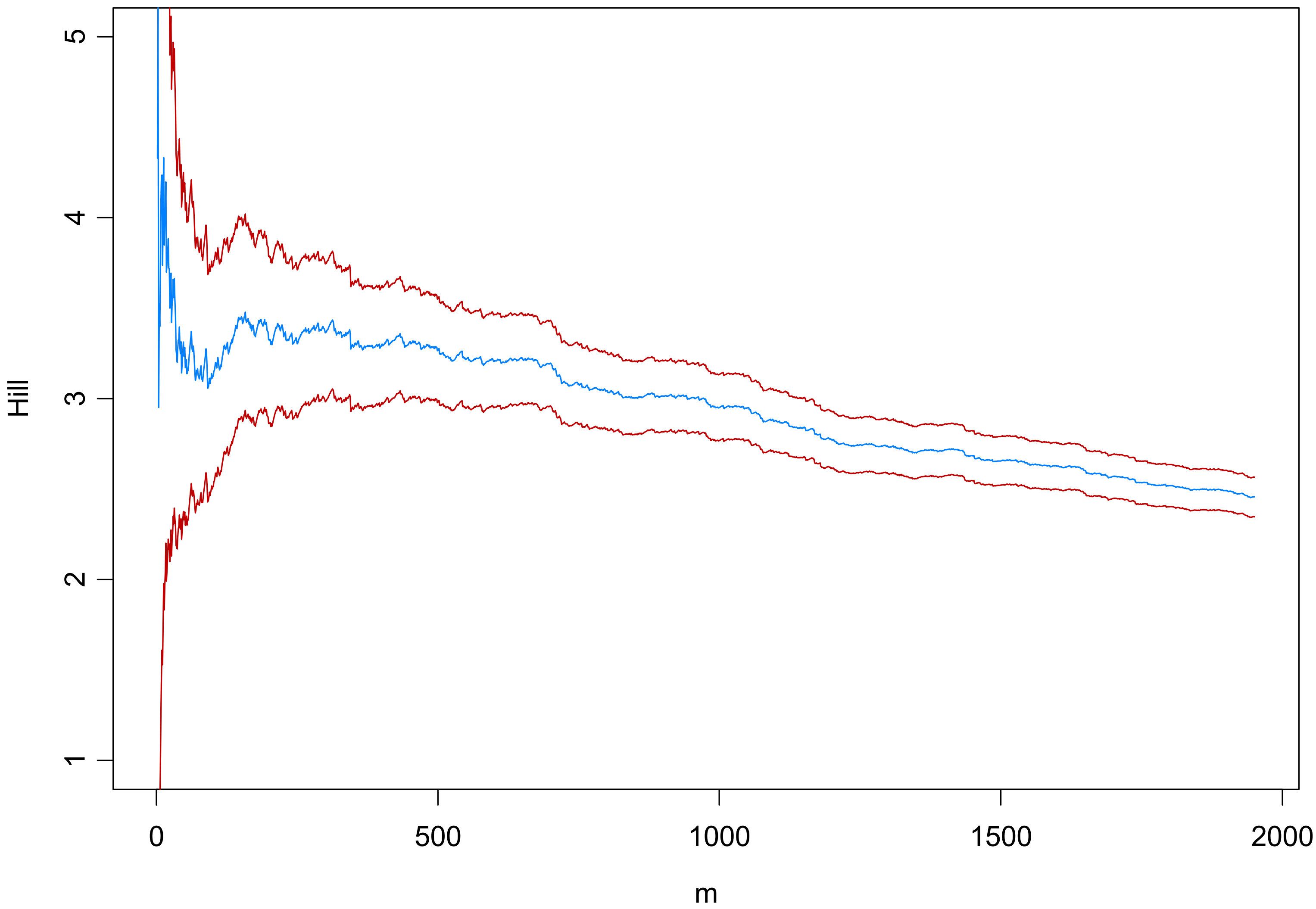
(See also GPD=generalized Pareto distribution.)

Hill's plot of tail index for IBM (1962-1981, 1982-2000)

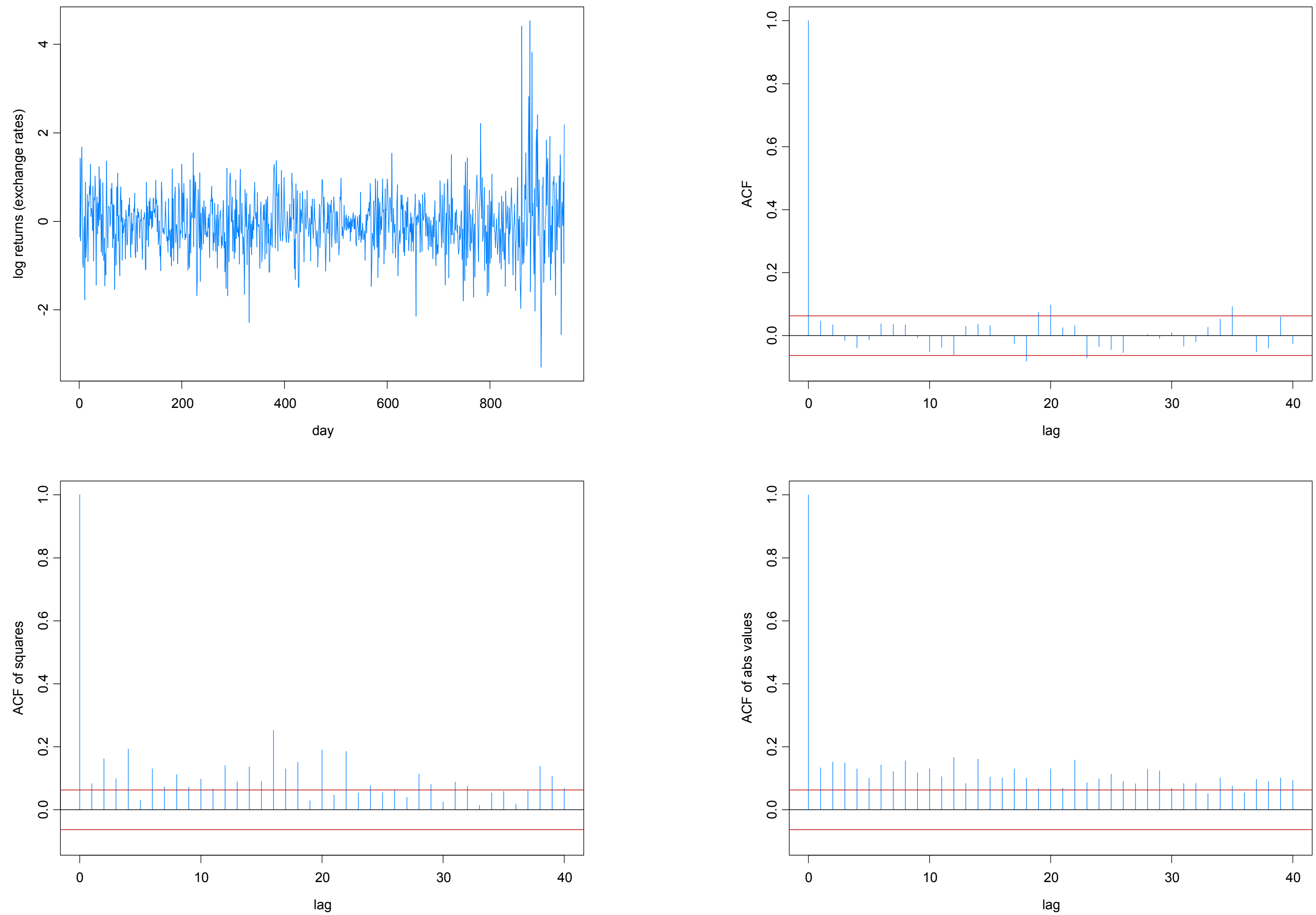


Example: S&P 500

Hill's estimate of alpha (Hill Horror plots-Resnick)

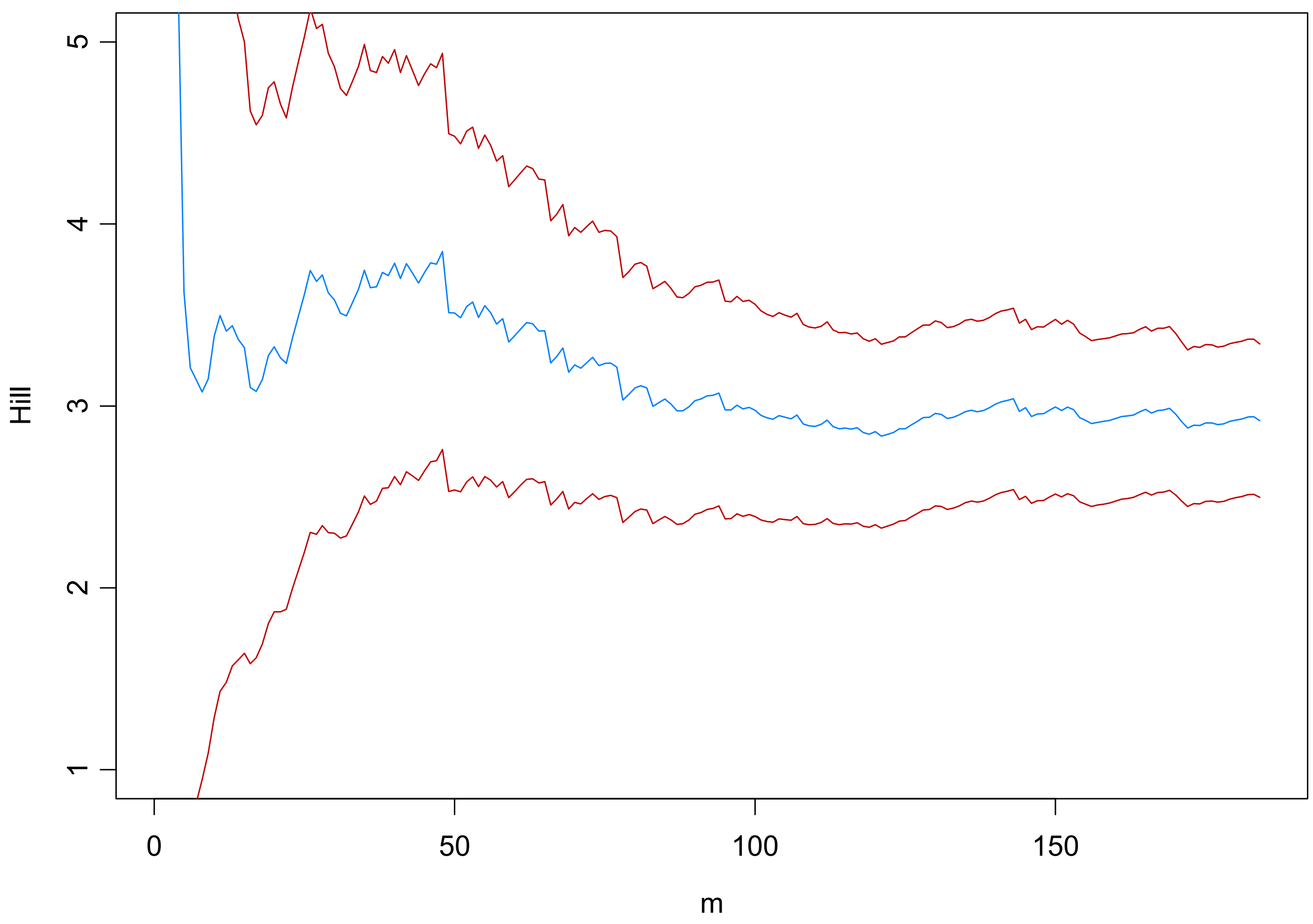


Example: Pound-Dollar Exchange Rates
(Oct 1, 1981 – Jun 28, 1985; Koopman website)



Example: Pound-Dollar Exchange Rates

Hill's estimate of alpha (Hill Horror plots-Resnick)



Multiplicative models for log(returns)

Basic model

$$X_t = \ln(P_t) - \ln(P_{t-1}) \quad \text{(log returns)}$$
$$= \sigma_t Z_t \,,$$

where

- $\{Z_t\}$ is IID with mean 0, variance 1 (if exists). (e.g. N(0,1) or a *t*-distribution with ν df.)
- $\{\sigma_t\}$ is the volatility process
- σ_t and Z_t are independent.

Properties:

- $EX_t = 0$, $\mathrm{Cov}(X_t, X_{t+h}) = 0$, $h>0$ (uncorrelated if $\mathrm{Var}(X_t) < \infty$)
- conditional heteroscedastic (condition on σ_t).

Multiplicative models for log(returns)-cont

$X_t = \sigma_t Z_t$ (observation eqn in state-space formulation)

Two classes of models for volatility:

(i) GARCH(p,q) process (General AutoRegressive Conditional Heteroscedastic-observation-driven specification)

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2 \ .$$

Special case: ARCH(1):

$$\begin{aligned} X_t^2 &= (\alpha_0 + \alpha_1 X_{t-1}^2) Z_t^2 \\ &= \alpha_1 Z_t^2 X_{t-1}^2 + \alpha_0 Z_t^2 \\ &= A_t X_{t-1}^2 + B_t \end{aligned}$$

(stochastic recurrence eqn)

$$\rho_{X^2}(h) = \alpha_1^h, \text{ if } \alpha_1^2 < 1/3.$$

Multiplicative models for log(returns)-cont

$X_t = \sigma_t Z_t$ (observation eqn in state-space formulation)

Two classes of models for volatility:

(i) GARCH(p,q) process (General AutoRegressive Conditional Heteroscedastic-observation-driven specification)

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2 \quad .$$

Special case: ARCH(1):

$$\begin{aligned} X_t^2 &= (\alpha_0 + \alpha_1 X_{t-1}^2) Z_t^2 \\ &= \alpha_1 Z_t^2 X_{t-1}^2 + \alpha_0 Z_t^2 \\ &= A_t X_{t-1}^2 + B_t \end{aligned}$$

(stochastic recurrence eqn)

$$\rho_{X^2}(h) = \alpha_1^h, \text{ if } \alpha_1^2 < 1/3.$$

Multiplicative models for log(returns)-cont

GARCH(2,1): $X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \alpha_2 X_{t-2}^2 + \beta_1 \sigma_{t-1}^2$.

Then $\mathbf{Y}_t = (\sigma_t^2, X_{t-1}^2)'$ follows the SRE given by

$$\begin{bmatrix} \sigma_t^2 \\ X_{t-1}^2 \end{bmatrix} = \begin{bmatrix} \alpha_1 Z_{t-1}^2 + \beta_1 & \alpha_2 \\ Z_{t-1}^2 & 0 \end{bmatrix} \begin{bmatrix} \sigma_{t-1}^2 \\ X_{t-2}^2 \end{bmatrix} + \begin{bmatrix} \alpha_0 \\ 0 \end{bmatrix}$$

Questions:

- Existence of a unique stationary solution to the SRE?
- Regular variation of the joint distributions?

Multiplicative models for log(returns)-cont

$X_t = \sigma_t Z_t$ (observation eqn in state-space formulation)

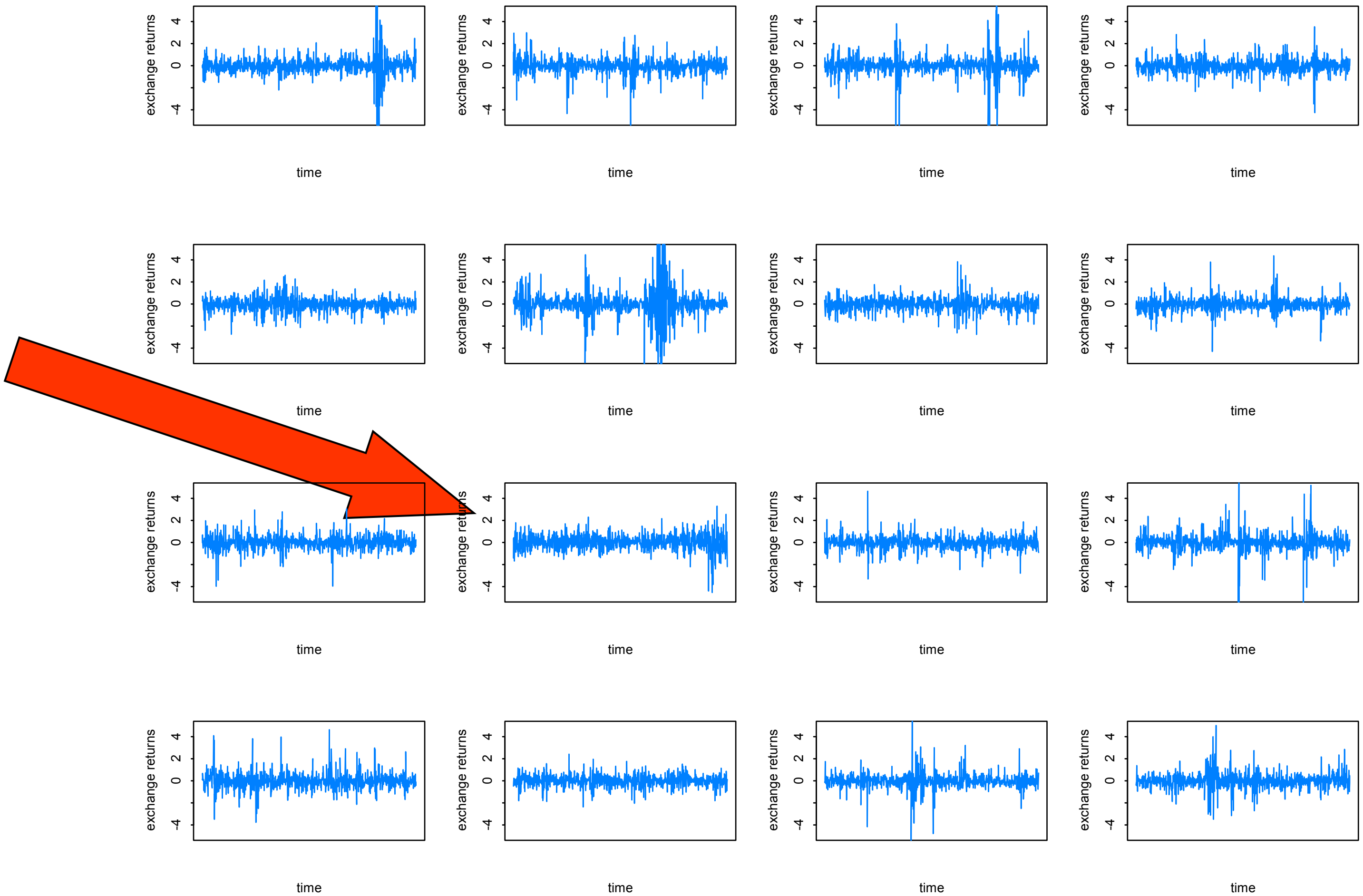
(ii) stochastic volatility process (parameter-driven specification)

$$\log \sigma_t^2 = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \quad \sum_{j=-\infty}^{\infty} \psi_j^2 < \infty, \{\varepsilon_t\} \sim \mathrm{IID\,N}(0,\sigma^2)$$

$$\rho_{X^2}(h) = Cor(\sigma_t^2, \sigma_{t+h}^2) / EZ_1^4$$

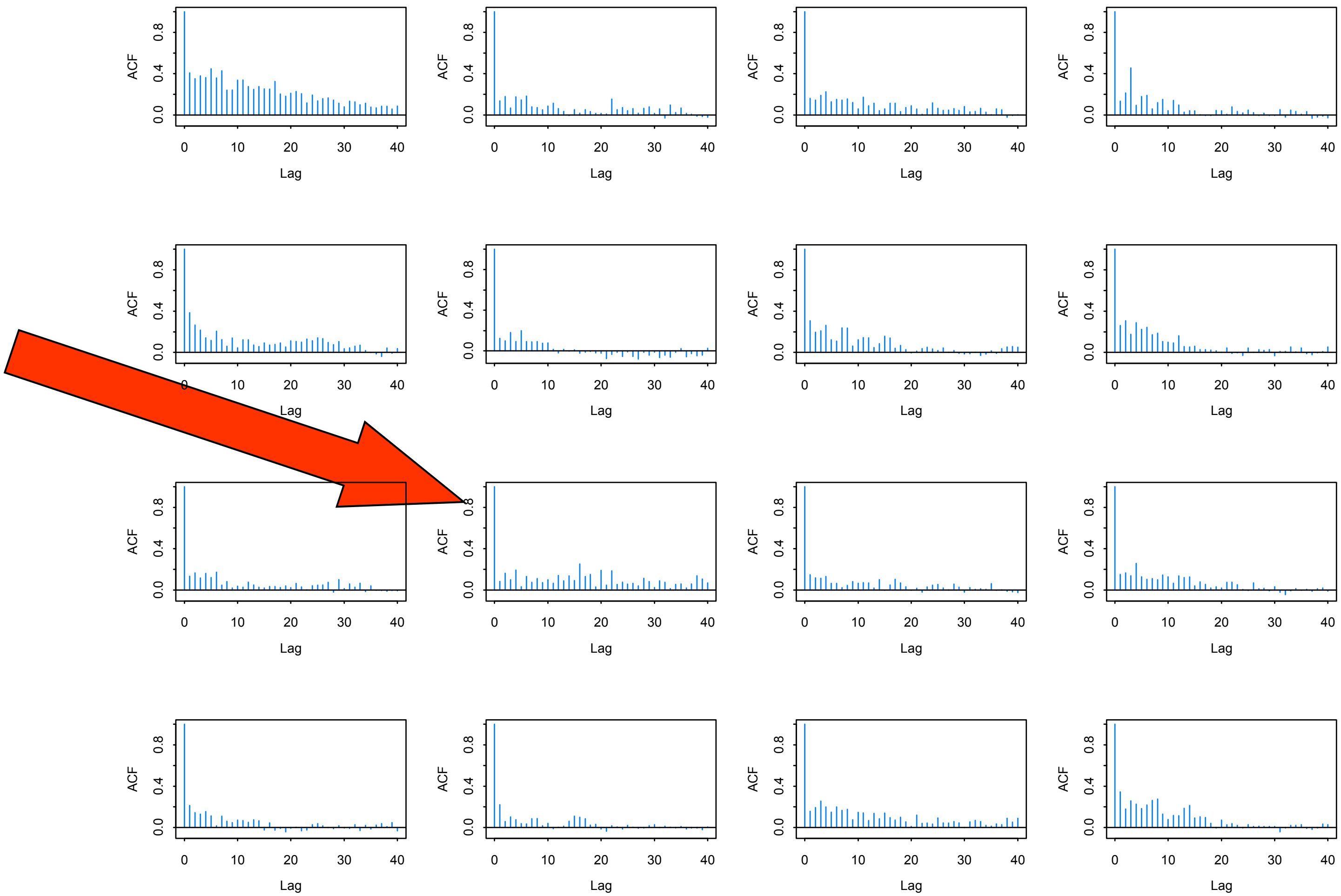
Stărică Plots for Pound-Dollar Exchange Rates

15 realizations from GARCH model fitted to exchange rates + exchange rate data. Which one is the real data?

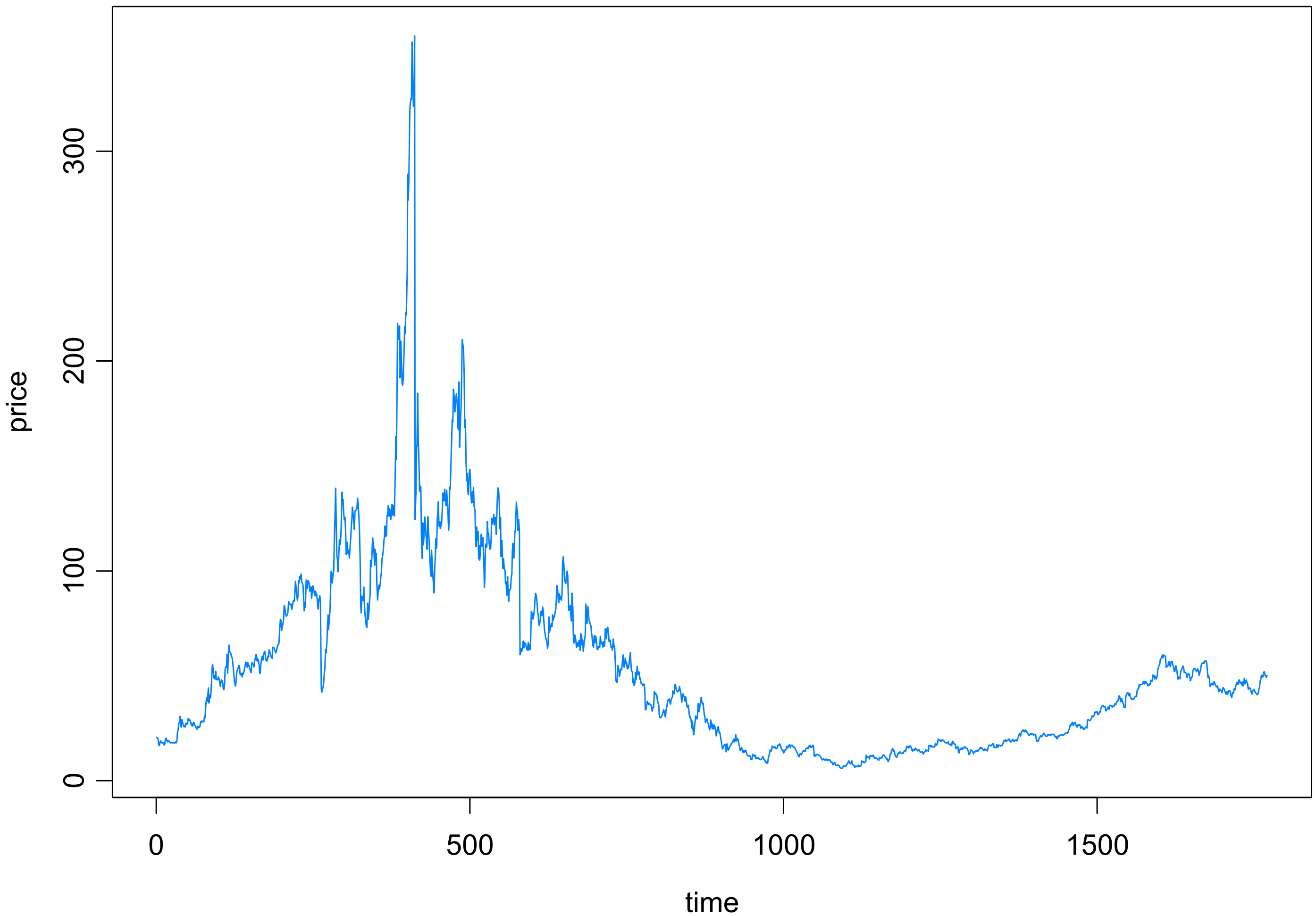


Stărică Plots for Pound-Dollar Exchange Rates

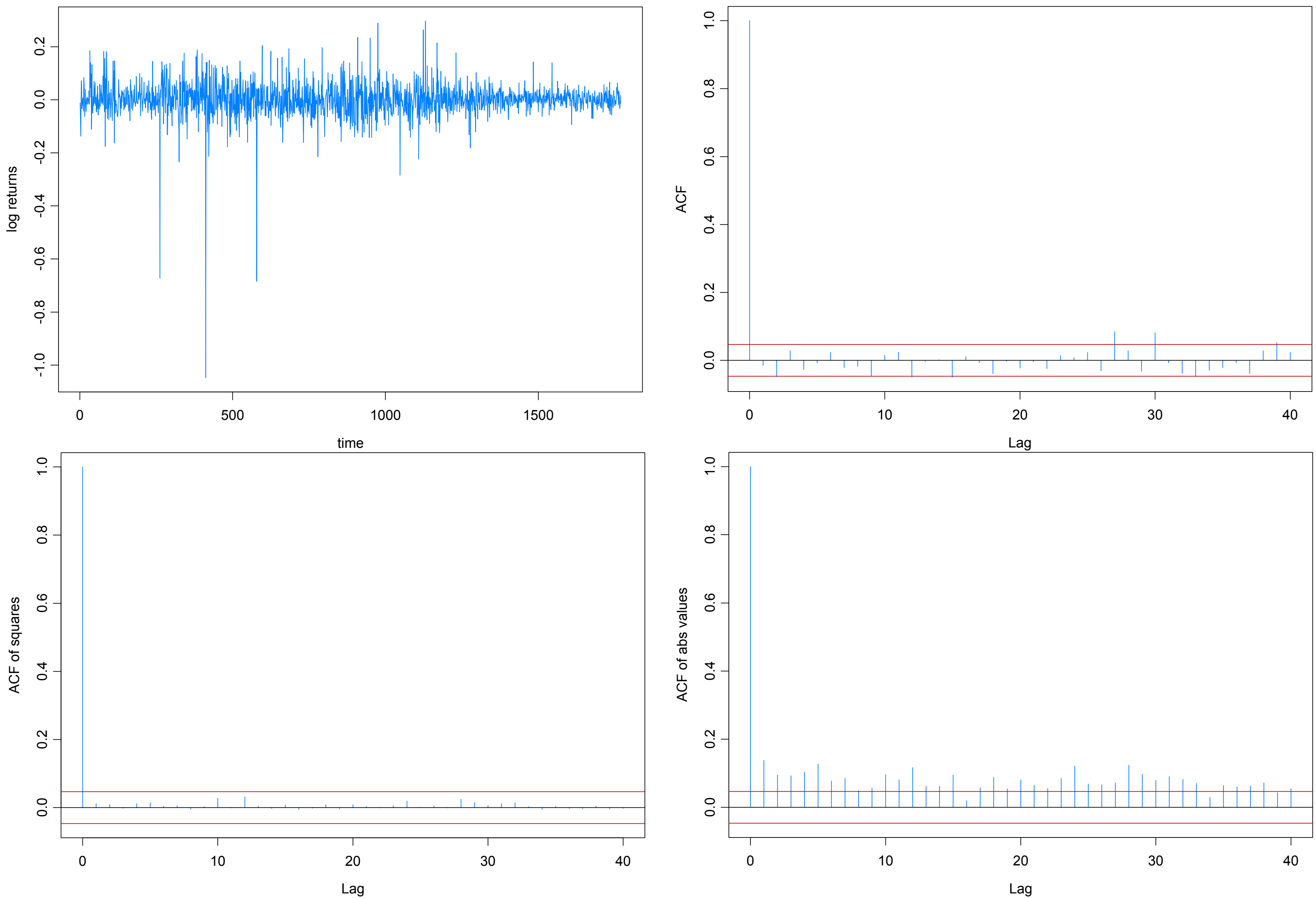
ACF of the squares from the 15 realizations from the GARCH model on previous slide.



Example: Amazon (May 16, 1997 – June 16, 2004)

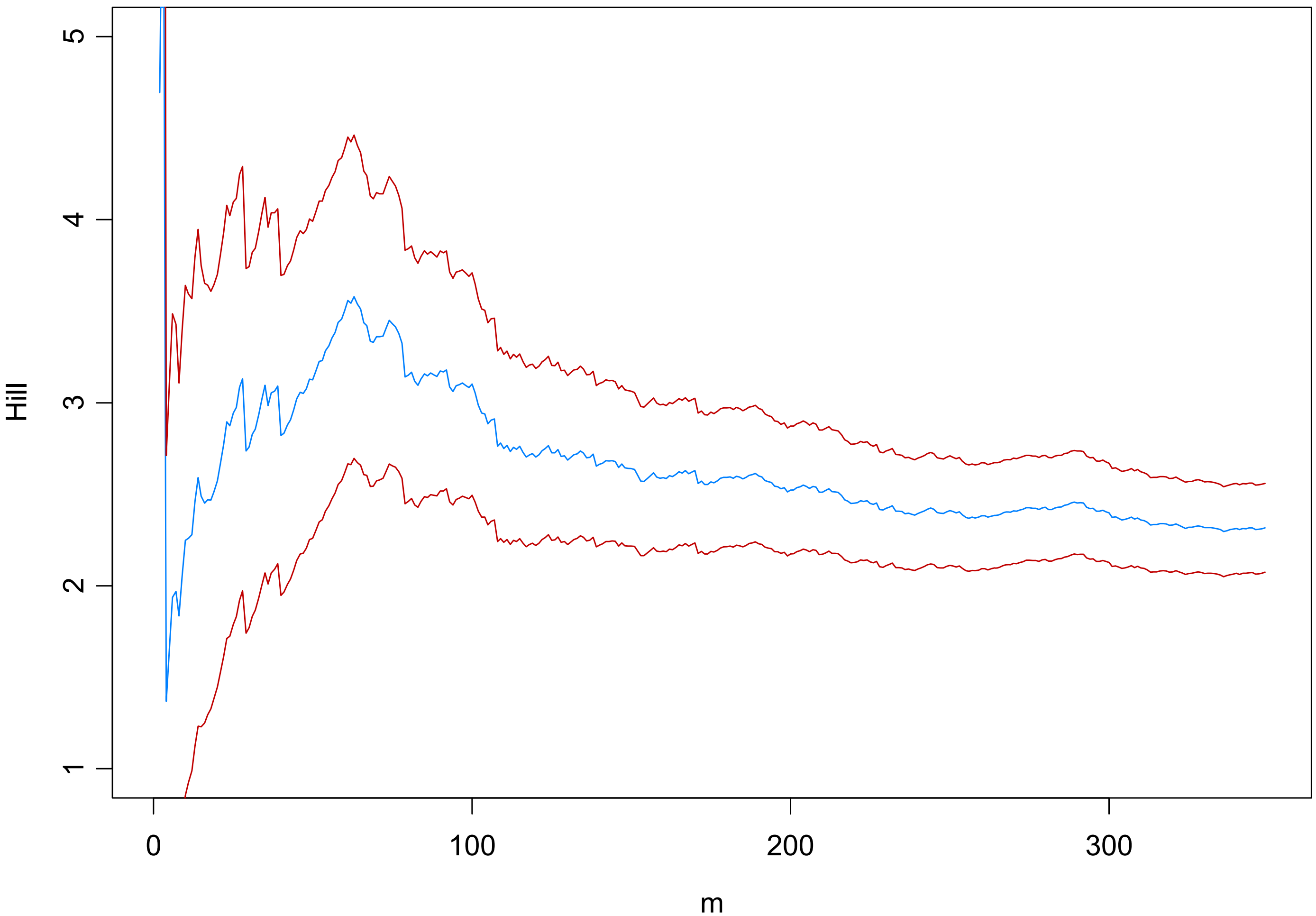


Example: Amazon-returns (May 16, 1997 – June 16, 2004)



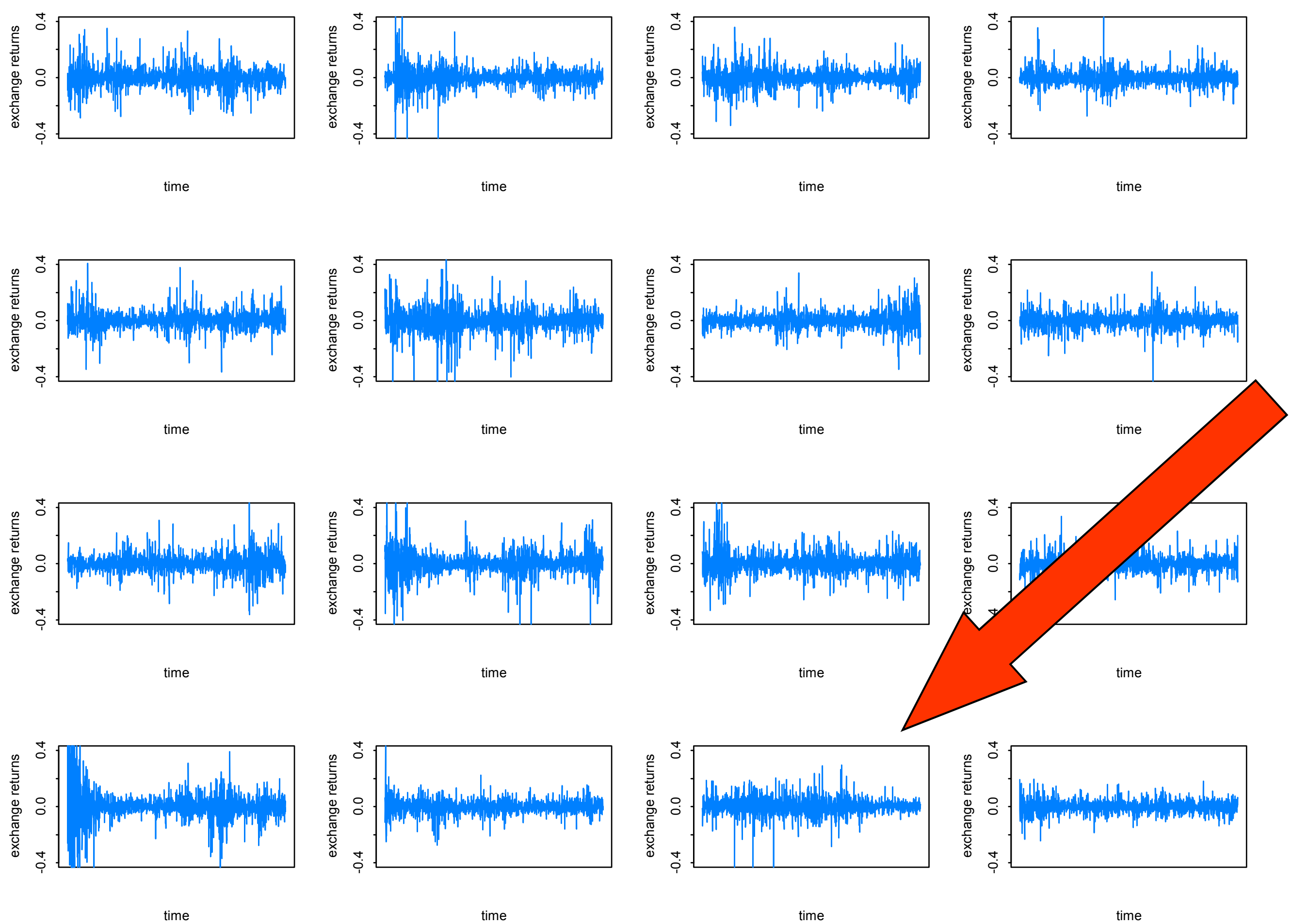
Example: Amazon-returns

Hill's estimate of alpha (Hill Horror plots-Resnick)



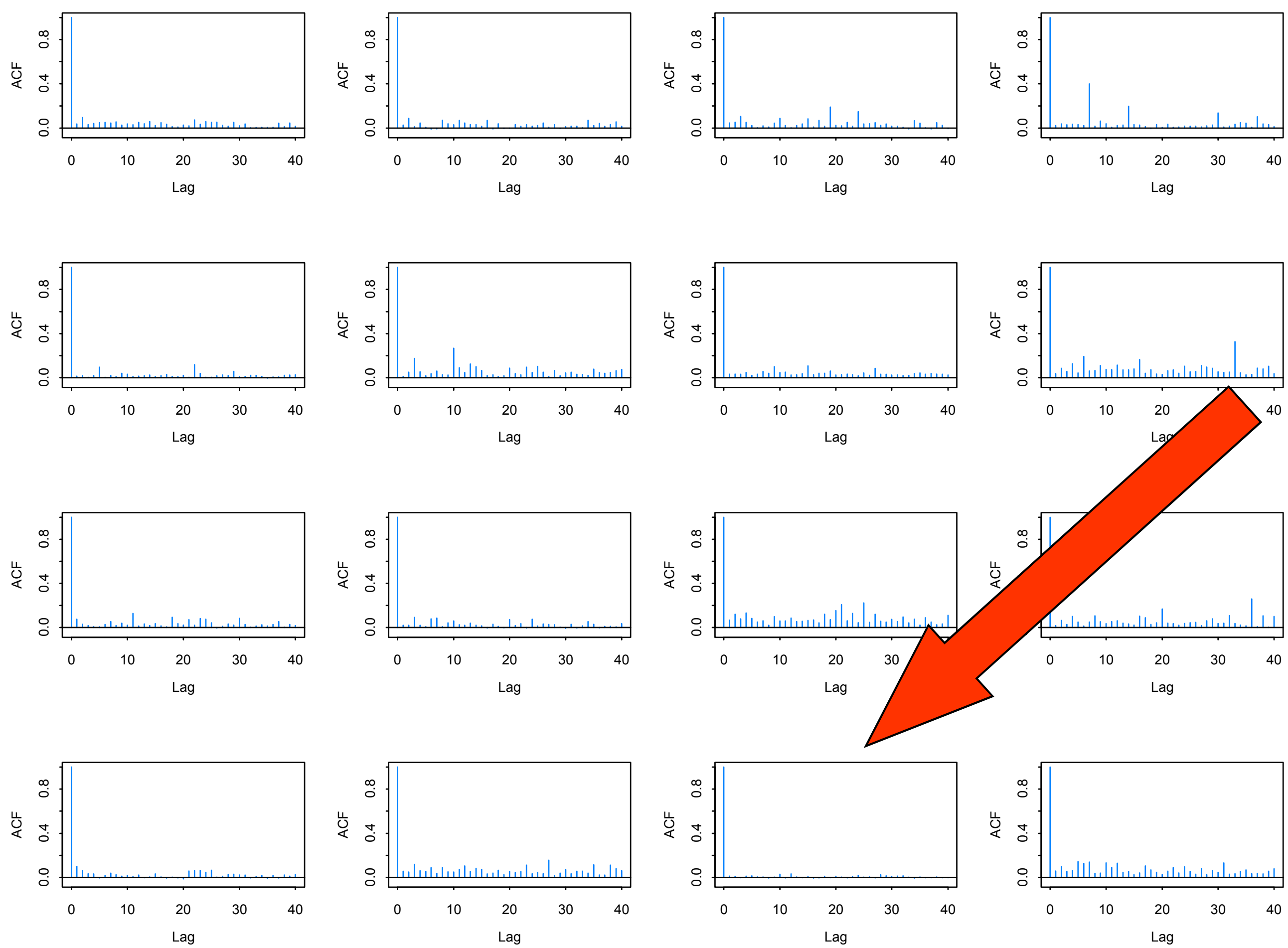
Stărică Plots for the Amazon Data

15 realizations from GARCH model fitted to Amazon + Amazon log-return data. Which one is the real data?



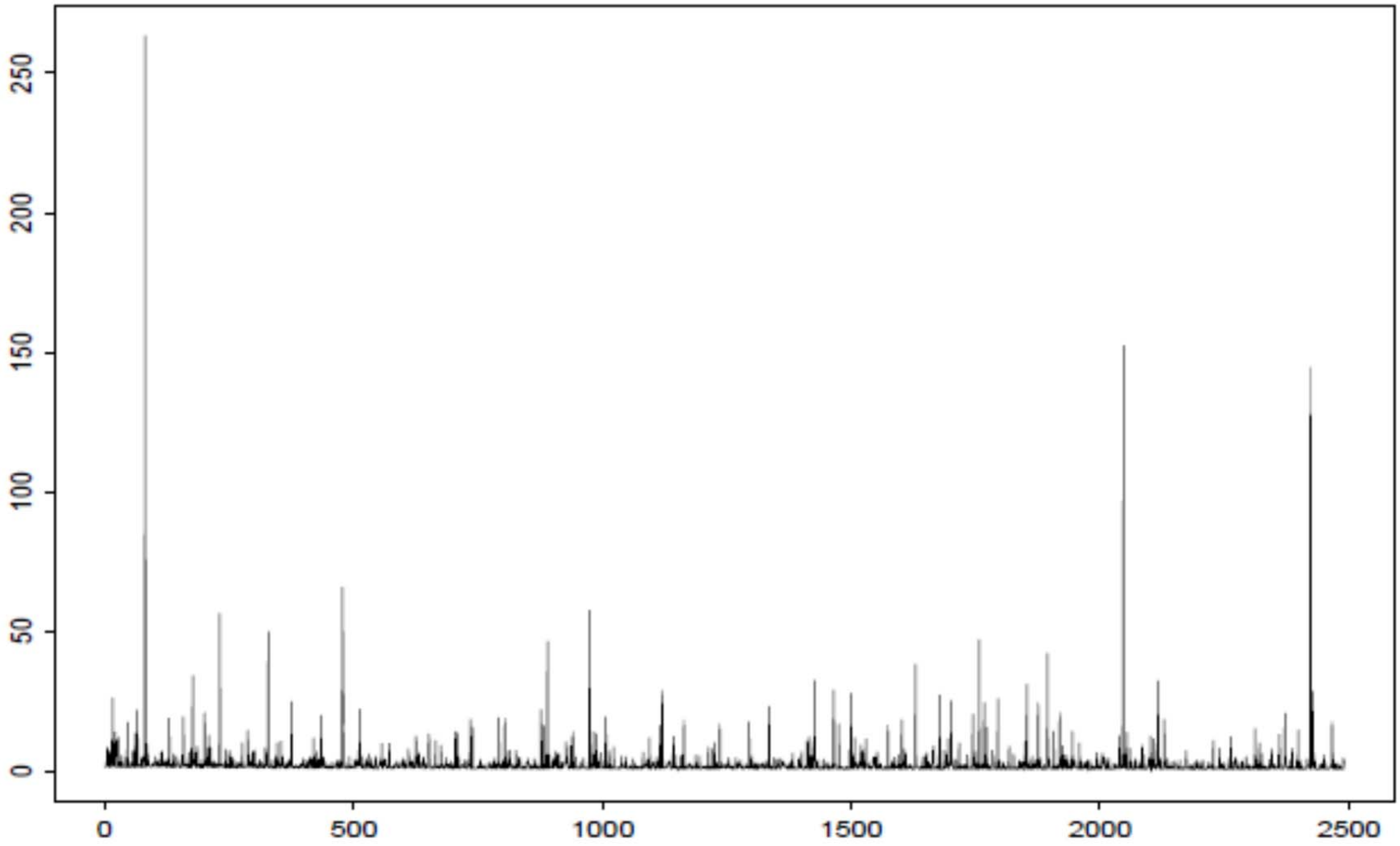
Stărică Plots for Amazon

ACF of the squares from the 15 realizations from the GARCH model on previous slide.



Heavy Tails in Insurance

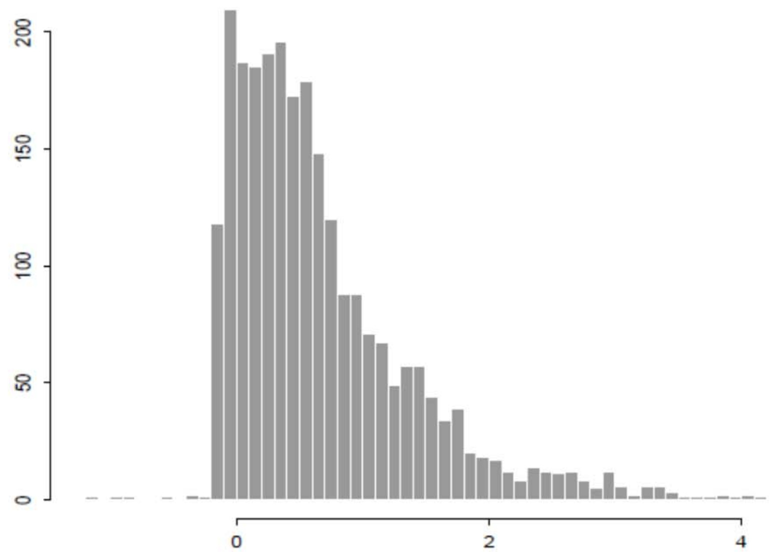
Danish fire insurance



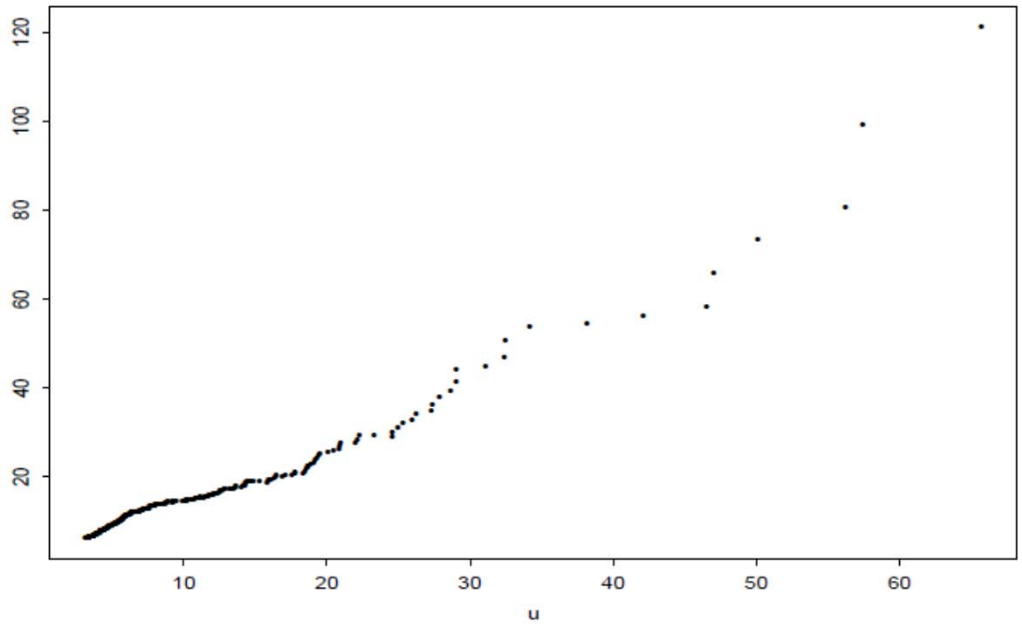
Heavy Tails in Insurance

Danish fire insurance

Histogram



Empirical mean excess function



Heavy-Tailed Distributions

Subexponential distributions: A positive rv X and its distribution F is said to be ***subexponential,*** if for iid copies of X and for any $n \geq 2$,

$$\frac{P(S_n > x)}{P(M_n > x)} \sim \frac{P(S_n > x)}{nP(X > x)} \to 1$$

as $x \to \infty$, where $S_n = X_1 + \cdots + X_n$ and $M_n = \max\{X_1, \ldots, X_n\}$.

Properties:

- Subexp distrs do not have moment generating functions.
- Examples include, log-normal, heavy-tailed Weibul, $F(x) = 1 - e^{-x^\tau}, x > 0$, for $\tau \in (0,1)$, and pareto distributions.
- Non-examples include: exponential, gamma, normal, finite support distrs.

Heavy-Tailed Distributions

Application to insurance: Suppose claim amounts X_i are iid with a subexponential distribution and claims occur according to a counting process $N(t)$, indep of the claim amounts. Then

$$P(S_{N(t)} > x) = \sum_{n=1}^{\infty} p_n P(S_n > x)\,,$$

where $p_n = P(N(t) = n)$. From the subexponentiality, we have

- $\frac{P(S_n > x)}{\bar{F}(x)} \leq K(1+\epsilon)^n$
- If $N(t)$ has a moment generation function in ngbhd of 0, $Ee^{hN(t)} < \infty$, then

$$\frac{P(S_{N(t)} > x)}{\bar{F}(x)} = \sum_{n=1}^{\infty} p_n \frac{P(S_n > x)}{\bar{F}(x)} \sim \sum_{n=1}^{\infty} n p_n = EN(t).$$

So the total claim amount of a portfolio at a high threshold satisfies

$$P\big(S_{N(t)} > x\big) \sim EN(t)\bar{F}(x)$$

Regularly Varying Distributions

Regularly varying random variable: A positive rv X is RV if

$$\bar{F}(x) = P(X > x) = L(x)x^{-\alpha},$$

where $L(x)$ is a slowly varying function at ∞, i.e.,

$$\frac{L(tx)}{L(t)} \to 1, \qquad as\ t \to \infty.$$

Properties of RV random variables:

- Subexponential (Feller, 1971)
- Examples include Pareto, log-gamma, non-normal stable distributions, Cauchy, student t, Fréchet.

Regularly Varying Distributions

Properties of RV random variables (cont):

Moments. $E|X|^p < \infty (= \infty)$ if $p < \alpha$ $(p > \alpha)$. $E|X|^\alpha$ could be finite or infinite depending on the slowly varying function L.

Maxima. Choose $a_n = F^{-1}(1 - \frac{1}{n})$ so that $n\bar{F}(a_n) \to 1$, and

$$n\bar{F}(a_n x) \sim \frac{\bar{F}(a_n x)}{\bar{F}(a_n)} \to x^{-\alpha}, as\; n \to \infty.$$

Setting $M_n = \max\{X_1, \ldots, X_n\}$, we have

$$P(a_n^{-1} M_n \le x) = F^n(a_n x) = \left(1 - \frac{n\bar{F}(a_n x)}{n}\right)^n \to \exp(-x^{-\alpha}).$$

The limit is called the **Fréchet** distribution.

Regularly Varying Distributions

Partial sums.

$$a_n^{-1}(S_n - b_n) \xrightarrow{d} S_\alpha, \qquad \alpha \in (0,2],$$

where S_α has a stable distribution with index α. ($\alpha = 2$ *is normal*, $\alpha = 1$ *is Cauchy*).

Joint convergence.

$$a_n^{-1}(M_n, S_n - b_n) \xrightarrow{d} (M_\alpha, S_\alpha)$$

where M_α and S_α have Frechet and stable distributions, respectively.

Self-normalized convergence.

$$(S_n - b_n)/M_n \xrightarrow{d} S_\alpha/M_\alpha$$

Regularly Varying Distributions

RV equivalences: $X > 0$ is RV iff for all x > 0,

$$\frac{P(X > tx)}{P(X > t)} \to x^{-\alpha} \; as \; t \to \infty.$$

Sequential definition. Replace t with a_n, where $P(X > a_n) \sim \frac{1}{n}$, then $X > 0$ is RV iff

$$nP(a_n^{-1}X > x) \to x^{-\alpha} \;\; as \; n \to \infty.$$

Vague convergence. $X > 0$ is RV iff

$$nP(a_n^{-1}X \in \cdot) \overset{v}{\to} \mu(\cdot) \;\; on \; (0, \infty]$$

where $\overset{v}{\to}$ denotes ***vague*** convergence, i.e.,

$$nP(a_n^{-1}X \in A) \to \mu(A)$$

for every Borel set A bounded away from 0 with $\mu(\partial A) = 0$.

Regularly Varying Distributions

Vague convergence. $X > 0$ is RV iff

$$nP(a_n^{-1}X \in \cdot) \xrightarrow{v} \mu(\cdot) \quad on\ (0, \infty]$$

The measure μ satisfies the homogeneity property

$$\mu(tA) = t^{-\alpha}\mu(A),$$

since

$$\mu(tA) = \lim_{n\to\infty} \frac{P(a_n^{-1}X \in tA)}{P(X > ta_n)} \frac{P(X > ta_n)}{P(X > a_n)} = \mu(A)t^{-\alpha}$$

Remark: The vague convergence equivalency allows one to easily extend to all of $\mathbb{R}$ and to the multivariate case $\mathbb{R}^d$.

Regularly Varying Distributions

Extension to $\overline{\mathbb{R}} \setminus \{0\}$: A rv X is RV iff $|X|$ is RV, i.e.,

$$\frac{P(|X| > tx)}{P(|X| > t)} \to x^{-\alpha} \; as \; t \to \infty$$

and the tail balancing condition (TBD) is met,

$$\frac{P(X > t)}{P(|X| > t)} \to p \quad and \quad \frac{P(X \leq t)}{P(|X| > t)} \to q \quad as \; t \to \infty,$$

where $p, q \geq 0, with \; p + q = 1.$

Sequential definition. Replace t with a_n, where $P(|X| > a_n) \sim \frac{1}{n}$, then

X is RV iff

$$nP(a_n^{-1}|X| > x) \to x^{-\alpha} \quad as \; n \to \infty$$

and the TBD is met, i.e.,

$$nP(a_n^{-1}X > x) \to px^{-\alpha} \quad and \quad nP(a_n^{-1}X \leq -x) \to q(-x)^{-\alpha}$$

Regularly Varying Distributions

Vague convergence: A rv X is RV iff

$$\frac{P(X \in t\,\cdot)}{P(|X| > t)} \to \mu(\cdot) \quad as\, t \to \infty$$

and if this is the case, then

$$\mu(dx) = p\alpha x^{-\alpha-1} 1_{(0,\infty]}(x)dx + q\alpha(-x)^{-\alpha-1} 1_{[-\infty,0)}(x)dx.$$

Regularly Varying Distributions

Examples:

- Pareto: $\bar{F}(x) = \frac{c^\alpha}{(x+c)^\alpha}, \quad x \geq 0, c > 0, \alpha > 0.$

- Student *t*: $f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left(1 + \frac{\mathrm{x}^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \ x \geq 0.$

- log-gamma: $\frac{\alpha^\beta}{\Gamma(\beta)}(\log x)^{\beta-1}x^{-\alpha-1}, \ x \geq 1, \alpha, \beta > 0.$

Result: Using Karmata's theorem, if a density f is regularly varying with index $-\alpha, \alpha > 1$, then

$$\bar{F}(x) = \int_x^\infty f(y)dy \sim (\alpha - 1)^{-1} x f(x)$$

is regularly varying with index $\alpha + 1$. Apply to above examples.

Point Process Methods

Dirac measure $\epsilon_x(\cdot)$: $\epsilon_x(A) = \begin{cases} 1, & if\ x \in A, \\ 0, & if\ x \notin A. \end{cases}$

Point process: $N_n(\cdot) = \sum_{t=1}^{n} \epsilon_{a_n^{-1} X_t}(\cdot)$, where (X_t) is an iid sequence of RV random variables.

Note: For A bounded away from 0,

$$N_n(A) \sim B(n, P(a_n^{-1} X \in A))$$

and since $EN_n(A) = nP(a_n^{-1} X_1 \in A) \to \mu(A)$,

$$N_n(A) \xrightarrow{d} N(A),$$

where $N(A) \sim Pois\big(\mu(A)\big)$. More generally,

$$N_n \xrightarrow{d} N,$$

where N is a Poisson process with intensity measure given by $\mu(dx)$.

Point Process Methods

$$N_n \xrightarrow{d} N$$

Point process convergence and applications: This convergence means that for any finite collection of disjoint sets $A_1, \dots, A_k$ that are *continuity* sets of μ and bounded away from zero,

$$\left(N_n(A_1), \dots, N_n(A_k)\right) \xrightarrow{d} \left(N(A_1), \dots, N(A_k)\right).$$

Convergence of extremes. Let $X_{(n)} \geq X_{(n-1)} \geq \cdots \geq X_{(1)}$ be the decreasing. order statistics. Since

$$\left\{a_n^{-1} X_{(n-k+1)} \leq x\right\} = \{N_n(x, \infty) \leq k-1\},$$

$$P\left(a_n^{-1} X_{(n)} \leq x_1, \dots, a_n^{-1} X_{(n-k+1)} \leq x_k\right)$$

$$= P(N_n(x_1, \infty) \leq 0, \dots, N_n(x_k, \infty) \leq k-1)$$

$$\rightarrow P(N(x_1, \infty) \leq 0, \dots, N(x_k, \infty) \leq k-1).$$

Point Process Methods

In particular, if $X \geq 0$,

$$P\left(a_n^{-1} X_{(n)} \leq x_1\right) \to P(N(x_1, \infty) \leq 0) = \exp(-\mu(x_1, \infty]) = \exp(-x_1^{-\alpha}).$$

$$P\left(a_n^{-1} X_{(n)} \leq x_1, a_n^{-1} X_{(n-1)} \leq x_2\right) \to P(N(x_1, \infty) \leq 0, N(x_2, \infty] \leq 1)$$

$$= \begin{cases} P(N(x_1, \infty) \leq 0), & if\ x_2 \geq x_1, \\ P(N(x_1, \infty) \leq 0) P(N(x_2, x_1] \leq 1), & if\ x_1 > x_2. \end{cases}$$

$$= \begin{cases} \exp(-x_1^{-\alpha}), & if\ x_2 \geq x_1, \\ \exp(-x_2^{-\alpha})\,(1 + x_2^{-\alpha} - x_1^{-\alpha}) & if\ x_1 > x_2. \end{cases}$$

Remark. Let (E_i) be an iid sequence of unit exponentials and set $\Gamma_i = E_1 + \cdots + E_i$. Then the Γ_i are the points of a homogeneous Poisson process, N^*. That is, $N^* = \sum_{i=1}^{\infty} \epsilon_{\Gamma_i}$.

Point Process Methods

$$N^* = \sum_{i=1}^{\infty} \epsilon_{\Gamma_i}$$

Now making the change of measure under the transformation

$$h: x \to x^{-1/\alpha},$$

it follows that $N := N^* \circ h^{-1}$ is a Poisson process with intensity measure

$$\mu(x,\infty] = \lambda \circ h^{-1}(x,\infty] = \lambda[0, x^{-\alpha}) = x^{-\alpha},$$

where λ is Lebesgue measure. It follows that $N = \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-1/\alpha}}$ and

$$N_n = \sum_{t=1}^{n} \epsilon_{a_n^{-1} X_t} \xrightarrow{d} \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-1/\alpha}}$$

Hence,

$$a_n^{-1}(X_{(n)}, \ldots, X_{(n-k+1)}) \xrightarrow{d} (\Gamma_1^{-1/\alpha}, \ldots, \Gamma_k^{-1/\alpha})$$

Point Process Methods

$$N_n = \sum_{t=1}^{n} \epsilon_{a_n^{-1} X_t} \xrightarrow{d} \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-1/\alpha}}$$

Convergence of spacings:

$$a_n^{-1}(X_{(n)} - X_{(n-1)}, \dots, X_{(n-k+1)} - X_{(n-k)}) \xrightarrow{d} (\Gamma_1^{-1/\alpha} - \Gamma_2^{-1/\alpha}, \dots, \Gamma_k^{-1/\alpha} - \Gamma_{k+1}^{-1/\alpha})$$

Heuristics for Hill's estimator: For *k* fixed,

$$(X_{(n)}, \dots, X_{(n-k+1)})/X\_((n-k)) \xrightarrow{d} (\Gamma_1^{-\frac{1}{\alpha}}, \dots, \Gamma_k^{-\frac{1}{\alpha}})/\Gamma_{k+1}^{-\frac{1}{\alpha}}$$

$$\frac{1}{k} \sum_{i=0}^{k-1} \left(\log(X_{(n-i)}) - \log(X_{(n-k)}) \right) \xrightarrow{d} \frac{1}{\alpha k} \sum_{i=0}^{k-1} (-\log \Gamma_i / \Gamma_{k+1})$$

Now the RHS is equal in distribution to $\frac{1}{\alpha k} \sum_{i=0}^{k-1} (-\log U_i) \triangleq \frac{1}{\alpha k} \sum_{i=0}^{k-1} E_i$,

where (U_i) is iid unit exponential.

Point Process Methods

So,

$$\frac{1}{k}\sum_{i=0}^{k-1}\left(\log(X_{(n-i)}) - \log(X_{(n-k)})\right) \xrightarrow{d} \frac{1}{\alpha k}\sum_{i=0}^{k-1} E_i$$

Letting $k \to \infty$, the RHS converges to α^{-1}. Subtracting α^{-1} and rescaling by $\sqrt{k}$, we obtain

$$\sqrt{k}\frac{1}{\alpha k}\sum_{i=0}^{k-1}(E_i - 1) \xrightarrow{d} N(0, \alpha^{-2}).$$

Denoting the LHS above as $\hat{\alpha}^{-1}$this suggests,

$$\sqrt{k}(\hat{\alpha}^{-1} - \alpha^{-1}) \xrightarrow{d} N(0, \alpha^{-2}).$$

(The rigorous proof takes a bit more work!)

Point Process Methods

Point process representation for general case $X \in \mathbb{R}$ is RV: For this case, let θ_i be an iid sequence such that

$$P(\theta_i = 1) = p \; and \; P(\theta_i = -1) = q = 1 - p.$$

Then

$$N_n = \sum_{t=1}^{n} \epsilon_{a_n^{-1} X_t} \xrightarrow{d} N = \sum_{i=1}^{\infty} \epsilon_{\theta_i \Gamma_i^{-1/\alpha}}$$

The limit point process is again a Poisson process with intensity measure $\mu(dx) = p\alpha x^{-\alpha-1} 1_{(0,\infty]}(x)dx + q\alpha x^{-\alpha-1} 1_{[-\infty,0)}(x)dx$.

So, provided $p, q > 0$, we have

$$P\left(a_n^{-1} X_{(n)} \le x\right) = P(N_n(x, \infty] = 0) \to P(N(x, \infty] = 0) = \exp(-p x^{-\alpha})$$

$$P\left(a_n^{-1} X_{(1)} > -x\right) = P(N_n[-\infty, -x] = 0) \to P(N_n[-\infty, -x] = 0)$$

$$= \exp(-q(-x)^{-\alpha})$$

Point Process Methods

Partial sum convergence: Suppose (X_t) is iid RV with $\alpha \in (0,2)$ and

set $S_n = X_1 + \cdots + S_n$. Recall $N_n = \sum_{t=1}^{n} \epsilon_{a_n^{-1} X_t} \xrightarrow{d} N = \sum_{i=1}^{\infty} \epsilon_{\theta_i \Gamma_i^{-1/\alpha}}$

If $\alpha \in (0,1)$, then

$$a_n^{-1} S_n \xrightarrow{d} S := \sum_{i=1}^{\infty} \theta_i \Gamma_i^{-\frac{1}{\alpha}} \quad \textit{limit has a stable distr}$$

If $\alpha \in (1,2)$, then

$$a_n^{-1}(S_n - b_n) \xrightarrow{d} S := \sum_{i=1}^{\infty} \left(\theta_i \Gamma_i^{-\frac{1}{\alpha}} - (p-q) E(\Gamma_i^{-\frac{1}{\alpha}} I(\Gamma_i^{-\frac{1}{\alpha}} \leq 1)) \right)$$

where $b_n = nE\big(X_1 I(|X_1| \leq a_n)\big)$.

Remark: The proofs require a bit more than a direct application of the continuous mapping theorem.

Time Series Examples

Linear time series models: Suppose (Z_t) is iid RV with index α and consider

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} ,$$

where the ψ_j are weights in a linear filter.

Remark: Often one assumes that the filter is *causal,* in which case the sum starts at 0. That is, X_t is a function of the past of the noise, $Z_s, s \leq t$.

For an AR(1) process, $X_t = \phi X_{t-1} + Z_t$, with $|\phi| < 1$, we have

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j} .$$

Time Series Examples

Assumptions on the filter weights: Need convergence of the infinite series $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$. Assume

$$\sum_{j=-\infty}^{\infty} \left|\psi_j\right|^{\delta} < \infty, \qquad \delta < \alpha, \delta \leq 1.$$

These conditions can be weakened (see Mikosch and Samorodnitsky (2000).

Relating tail of X_t to that of the noise Z_t $(M\&S, 2000)$:

$$\frac{P(X_t > x)}{P(|Z| > x)} \to \sum_{j=-\infty}^{\infty} \left(p(\psi_j^+)^{\alpha} + q(\psi_j^-)^{\alpha}\right) \;\; and \;\; \frac{P(|X_t| > x)}{P(|Z| > x)} \to \sum_{j=-\infty}^{\infty} \left|\psi_j\right|^{\alpha}.$$

where $\psi_j^+ = \max(0, \psi_j), \psi_j^- = \max(0, -\psi_j)$.

It follows that X_t is also RV with the same index as the noise Z_t.

Time Series Examples

Proof of this result: This really follows by the subexponentiality of RV random variables. That is, for non-negative Z_t,

$$\frac{P(Z_1 + \cdots + Z_n > x)}{P(Z_1 > x)} \to n \quad as\ x \to \infty.$$

Adding non-negative coefficients in front of the $Z_t's$ doesn't change much, i.e., for $\psi_j > 0$,

$$\frac{P(\psi_1 Z_1 + \cdots + \psi_n Z_n > x)}{P(Z_1 > x)} \to \psi_1^\alpha + \cdots + \psi_n^\alpha \quad as\ x \to \infty.$$

The final step is to extend to $n \to \infty$; the case of possibly negative valued random variables and coefficients is more of a book-keeping problem.

Time Series Examples

Stochastic volatility models (SVM): These models are used for modeling log-returns.

Multiplicative models for log-returns

$$\begin{aligned} X_t &= \ln(P_t) - \ln(P_{t-1}), \quad \text{log-returns} \\ &= \sigma_t Z_t \end{aligned}$$

where

- $\{Z_t\}$ is IID with mean 0, variance 1 (if exists). (e.g. N(0,1) or a t-distribution with ν df.)
- $\{\sigma_t\}$ is the volatility process
- σ_t and Z_t are independent.

Properties:

- $EX_t = 0, Cov(X_t, X_{t+h}) = 0, h > 0$ (uncorrelated if $Var(X_t) < \infty$)
- conditional heteroscedastic (condition on σ_t).

Time Series Examples

Stochastic volatility models (SVM):

$$X_t = \sigma_t Z_t$$

where

- volatility process (σ_t) is a sequence of positive random variables
- noise sequence (Z_t) is iid and independent of (σ_t)

Assumptions used here:

- log-volatility process, $\log \sigma_t$ is a Gaussian linear time series process,

$$\log \sigma_t = \sum_{j=0}^{\infty} \psi_j \eta_{t-j}, \qquad (\eta_t) \sim IID\ N\left(0, \sigma_\eta^2\right).$$

Covariance function of $\mathrm{Y_t} = \log \sigma_t$:

$$\gamma_Y(h) = Cov(Y_t, Y_{t+h}) = \sum_{j=0}^{\infty} \psi_j \psi_{j+h} \sigma_\eta^2$$

Time Series Examples

Stochastic volatility models (SVM):

$$X_t = \sigma_t Z_t$$

- $(Z_t) \sim IID$ RV with index α (mean 0 and variance 1) if $\alpha > 2$.

If $\alpha > 2$, then

$$Cov(X_0, X_h) = 0, h \neq 0.$$

This is what we observe empirically, i.e., log-returns appear uncorrelated. On the other hand,

$$\gamma_{|X|^p}(h) = cov(|X_0|^p, |X_h|^p)$$
$$= E\left(\sigma_0^p \sigma_h^p\right)\left(E(|Z_0|^p)\right)^2 - (E\sigma_0^p E(|Z_0|^p)^2$$
$$= \left(E(|Z_0|^p)\right)^2 cov(\sigma_0^p, \sigma_h^p)$$

Thus, the covariance of higher order moments of the log-returns can be modeling in a flexible way through the covariance function of the volatility process. In particular,

$$Cov\left(\sigma_0^p, \sigma_h^p\right) = E(e^{pY_0 + pY_h}) - (Ee^{pY_0})^2$$
$$= \exp(p^2 \gamma_Y(0))(\exp\left(\gamma_Y(h)\right) - 1)$$
$$\sim \text{const}\, \gamma_Y(h) \quad as\ h \to \infty.$$

Time Series Examples

In particular,

$$Cov(\sigma_0^p, \sigma_h^p) = E(e^{pY_0 + pY_h}) - (Ee^{pY_0})^2$$
$$= \exp(p^2 \gamma_Y(0))(\exp(\gamma_Y(h)) - 1)$$
$$\sim \text{const}\, \gamma_Y(h) \quad as\ h \to \infty,$$

so that

$$\gamma_{|X|^p}(h) \sim const\, \gamma_{\sigma^p}(h) \sim const\, \gamma_Y(h) \quad as\ h \to \infty.$$

Does X_t have RV tails if Z_t does? Answer is yes, which follows from Breiman's lemma.

Breiman Lemma: Suppose $Z > 0$ is RV with index α and $\sigma > 0$ is a random variable independent of Z with $E\sigma^\beta < \infty$ for some $\beta > \alpha$. Then

$$\lim_{x\to\infty} \frac{P(X > x)}{P(Z > x)} \to E\sigma^\alpha,$$

where $X = \sigma Z$.

Clearly, this result applies for the SVM, i.e., X_t is RV of the same order as the noise (similar to the linear time series).

Time Series Examples

Generalized **A**utoregressive **C**onditional **H**eteroscedastic (GARCH(p,q)):

$$X_t = \sigma_t Z_t$$

- $(Z_t) \sim IID$ (mean 0 and variance 1) if $var(Z) < \infty$.
- $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2,\ \alpha_i > 0, \beta_i > 0.$
- The most commonly used model is the GARCH(1,1) and we will focus on this one:

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Note that the volatility process satisfies a ***stochastic recurrence equation (SRE)***, which has the form:

$$\begin{aligned} \sigma_t^2 &= (\alpha_1 Z_{t-1}^2 + \beta_1)\sigma_{t-1}^2 + \alpha_0 \\ &= A_t \sigma_{t-1}^2 + B_t \end{aligned}$$

where $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$ and $B_t = \alpha_0$. The general 1-dim'l SRE is

$$X_t = A_t X_{t-1} + B_t$$

Time Series Examples

$$X_t = A_t X_{t-1} + B_t$$

Iterating backwards,

$$
\begin{aligned}
X_t &= A_t X_{t-1} + B_t \\
&= A_t A_{t-1} X_{t-2} + A_t B_{t-1} + B_t \\
&= A_t A_{t-1} A_{t-2} X_{t-3} + A_t A_{t-1} B_{t-2} + A_t B_{t-1} + B_t \\
&\vdots \\
&= \sum_{i=1}^{\infty} A_t A_{t-1} \cdots A_{t-i+2} B_{t-i+1}
\end{aligned}
$$

where $A_t A_{t-1} \cdots A_{t-k} X_{t-k-1} \rightarrow 0$ in prob. Note that the infinite sum is a solution to the SRE provided the sum exists. (See next result.)

Time Series Examples

$$X_t = A_t X_{t-1} + B_t$$

Proposition. Let (A_t, B_t) be an iid sequence with $A_t \geq 0$ $a.s.$ Assume one of the following conditions hold:

1. $P(A = 0) > 0$
2. $P(A = 0) = 0, -\infty \leq E \log A < 0$ and $E \log^+ |B| < \infty$.

Then there exists a unique causal soln to the SRE given by the infinite series $X_t = \sum_{i=1}^{\infty} A_t A_{t-1} \cdots A_{t-i+2} B_{t-i+1}$.

GARCH(1,1): $A_t = \alpha_1 Z_t^2 + \beta_1, B_t = \alpha_0$. Unique stationary soln if

$$E(\log(\alpha_1 Z^2 + \beta_1)) < 0 \text{ } and \text{ } \alpha_0 > 0.$$

A sufficient condition for the former is

$$E(\log(\alpha_1 Z^2 + \beta_1)) \leq \log E(\alpha_1 Z^2 + \beta_1) = \log(\alpha_1 + \beta_1) < 0$$

if $\alpha_1 + \beta_1 < 1$.

Time Series Examples

If $\alpha_1 + \beta_1 < 1 \; and \; \alpha_0 > 0,$ then $Var(X_t) = \frac{\alpha_0}{1-\alpha_1-\beta_1}$.

However, it is possible to have a strictly stationary solution even when $\alpha_1 + \beta_1 \geq 1$, but in this case, $Var(X_t) = \infty$.

Theorem (Kesten,1973 and Goldie, 1991). Let $A > 0, B \geq 0,$ be rvs st.

- $\exists\, \alpha > 0 \; st. \, EA^{\alpha} = 1, E(A^{\alpha}\log^{+} A) < \infty, EB^{\alpha} < \infty$.
- Support of $log\ A$ is not concentrated on a lattice.
- $\forall\, x > 0, P(Ax + B = x) < 1$.

Then $\exists\, rv\ X$, independent of $A \; and \; B$, such that $X = AX$+B and

$$P(X > x) \sim c_{+} x^{-\alpha}, \qquad (asymptotic\ Pareto!)$$

where $c_{+} = E((AX + B^{\alpha}) - (AX)^{\alpha})/(\alpha E(A^{\alpha}\log A))$.

Time Series Examples

Comments:

- If A has a density wrt Lebesgue measure, then 1 and 2 are met.
- The function $h(r) = Ee^{r \log A}$ is strictly convex ($h''(r) > 0$) and since $h(0) = 1,\ h(r) < 1\ for\ small\ r$. Need $h(r) \geq 1\ for\ large\ r$ in which case $\exists\ \alpha$ s.t. $h(\alpha) = 1$.

GARCH(1,1): $A_t = \alpha_1 Z_t^2 + \beta_1$ so α satisfies

$$E(\alpha_1 Z_t^2 + \beta_1)^\alpha = 1.$$

This is easy to check when Z has a density function (like normal or t.)

Note: if $\alpha_1 + \beta_1 = 1$, then the GARCH(1,1) is called an ***INGARCH(1,1).*** In this case, $\boldsymbol{\alpha} = \mathbf{1}$ regardless of the distribution for *Z*, since

$$E(Z^2) = 1.$$

Time Series Examples

GARCH(1,1): So under mild assumptions on the noise distribution, we have σ_t^2 is RV, i.e.,

$$P(\sigma_t^2 > x) \sim c_+ x^{-\alpha} \Longrightarrow P(\sigma_t > x) \sim c_+ x^{-2\alpha}.$$

But $X_t = \sigma_t Z_t,$ with $Z_t \ and \ \sigma_t$ independent, it follows from Breiman's Lemma (with roles of $Z_t \ and \ \sigma_t$ interchanged) that

$$P(X_t > x) \sim EZ_+^{2\alpha} P(\sigma_t > x) \sim EZ_+^{2\alpha} c_+ \, x^{-2\alpha}$$

$$P(X_t \leq -x) \sim EZ_-^{2\alpha} P(\sigma_t > x) \sim EZ_-^{2\alpha} c_+ x^{-2\alpha},$$

where $Z_\pm$ is the $\pm$ part of Z.

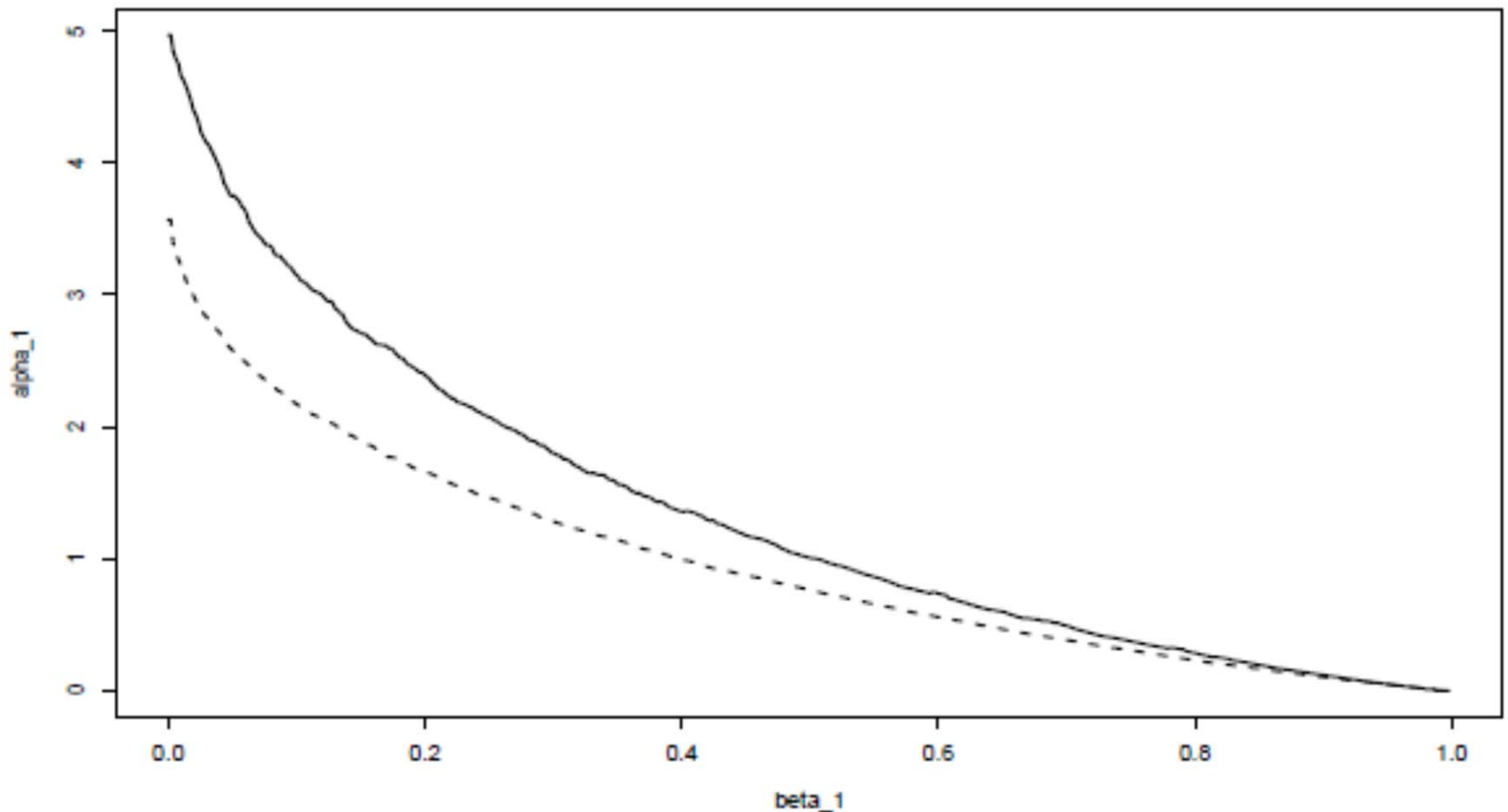
Summary: For GARCH(1,1) process (X_t) the marginal distribution is RV with index 2α, where

$$E(\alpha_1 Z_t^2 + \beta_1)^\alpha = 1.$$

Note: Breiman works here with just assuming $E|Z_t|^{2\alpha} < \infty$ since the tails of σ_t are asymptotic Pareto(2α).

Time Series Examples

GARCH(1,1): Stationary region for parameters (α_1, β_1) for normal distribution (below dotted) and t with 4 df (below solid). Region calculated via simulation 1,000,000 reps.



Time Series Examples

GARCH(1,1): $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2,\ \alpha_1 = .10$

$$P(X_t > x) \sim const\ x^{-2\alpha}$$

Normal distribution

β	.90	.80	.70	.60	.50	.40	.30	.20	.10	.00
2α	2.0	12.5	16.2	18.5	20.2	21.7	23.0	24.2	25.4	26.5

t- distribution, normalized to have variance 1

β	.90	.80	.70	.60	.50	.40	.30	.20	.10	.00
2α	2.0	3.68	3.83	3.88	3.91	3.92	3.93	3.93	3.94	3.94

Note: lighter tails $\Leftrightarrow$ shorter memory (smaller β)

Multivariate regular variation — univariate revisited

Def: The random variable *X* is *regularly varying with index* α if

$$P(|X|> t\,x)/P(|X|>t) \to x^{-\alpha} \text{ and } P(X> t)/P(|X|>t) \to p,$$

or, equivalently, if

$$P(X> t\,x)/P(|X|>t) \to px^{-\alpha} \text{ and } P(X< -t\,x)/P(|X|>t) \to qx^{-\alpha},$$

where $0 \le p \le 1$ and $p+q=1$.

Equivalence:

X is RV(α) *if and only if* $P(X \in t\,\bullet\,)/P(|X|>t) \to_v \mu(\bullet\,)$

($\to_v$ vague convergence of measures on R\{0}). In this case,

$$\mu(dx) = \left(p\alpha\, x^{-\alpha-1}\, I(x>0) + q\alpha\, (-x)^{-\alpha-1}\, I(x<0)\right) dx$$

Note: $\mu(tA) = t^{-\alpha}\,\mu(A)$ for every t and A bounded away from 0.

Multivariate regular variation — univariate revisited

Another formulation (polar coordinates):

Define the ± 1 valued rv Θ, $P(\Theta = 1) = p, P(\Theta = -1) = 1 - p = q$.

Then

X is RV(α) *if and only if*

$$\frac{P(\frac{X}{|X|} \in S, |X| > tx)}{P(|X| > t)} \to x^{-\alpha} P(\Theta \in S).$$

or

$$\frac{P(\frac{X}{|X|} \in \cdot, |X| > tx)}{P(|X| > t)} \stackrel{v}{\to} x^{-\alpha} P(\Theta \in \cdot).$$

($\stackrel{v}{\to}$ denotes vague convergence of measures on $S^0 = \{-1,1\}$).

Regular variation — multivariate case

Multivariate regular variation of $X = (X_1, \ldots, X_m)$: There exists a random vector $\Theta \in S^{m-1}$ such that

$$\frac{P(\frac{X}{|X|} \in \cdot, |X| > tx)}{P(|X| > t)} \stackrel{v}{\rightarrow} x^{-\alpha} P(\Theta \in \cdot).$$

($\stackrel{v}{\rightarrow}$ denotes vague convergence on S^{m-1}, unit sphere in R^m) .

- $P(\Theta \in \cdot)$. is called the spectral measure
- α is the index of X.

Equivalence:

$$\frac{P(X \in t \cdot)}{P(|X| > t)} \stackrel{v}{\rightarrow} \mu(\cdot)$$

μ is a measure on R^m which satisfies for $x > 0$ and A bounded away from 0, $\mu(xA) = x^{-\alpha} \mu(A)$.

Regular variation — multivariate case

A more intuitive equivalence: $X = (X_1, \ldots, X_m)$ is RV iff

$$\frac{P(|X| > tx)}{P(|X| > t)} \to x^{-\alpha}, \qquad as\ t \to \infty$$

and

$$P\left(\frac{X}{|X|} \in \cdot \,\middle|\, |X| > x\right) \xrightarrow{w} P(\Theta \in \cdot).$$

These 2 conditions correspond to :

- radial part of X is RV (univariate)
- angular part if independent of radial part when radial part is large.

Regular variation — multivariate case

Example: Let $X = R\Theta$, where $R > 0$ is RV with index α and independent of $\Theta \in S^{m-1}$. Then X is RV, since $|X| = R$ and $X/|X| = \Theta$. Hence

$$\frac{P(|X| > tx)}{P(|X| > t)} = \frac{P(|R| > tx)}{P(|R| > t)} \to x^{-\alpha}, \qquad as\ t \to \infty$$

and

$$P\left(\frac{X}{|X|} \in \cdot \,\middle|\, |X| > x\right) = P(\Theta \in \cdot \,||R| > x) = P(\Theta \in \cdot) \xrightarrow{w} P(\Theta \in \cdot).$$

If Θ has a uniform distribution, X has a ***spherical*** distribution.

Regular variation — multivariate case (cont)

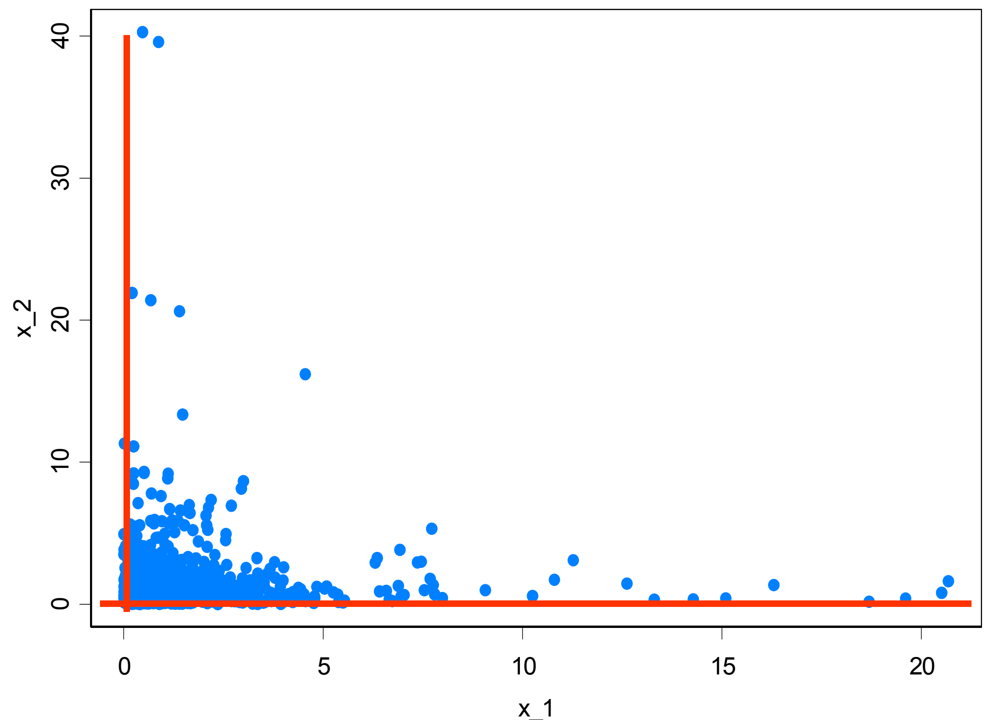
Examples:

1. If $X_1 > 0$ and $X_2 > 0$ are iid RV(α), then $\mathbf{X} = (X_1, X_2)$ is multivariate regularly varying with index α and *spectral distribution*

$$P(\theta = (0,1)) = P(\theta = (1,0)) = .5 \quad \text{(mass on axes).}$$

Interpretation: Unlikely that X_1 and X_2 are very large at the same time.

Figure: plot of (X_{t1}, X_{t2}) for realization of 10,000.



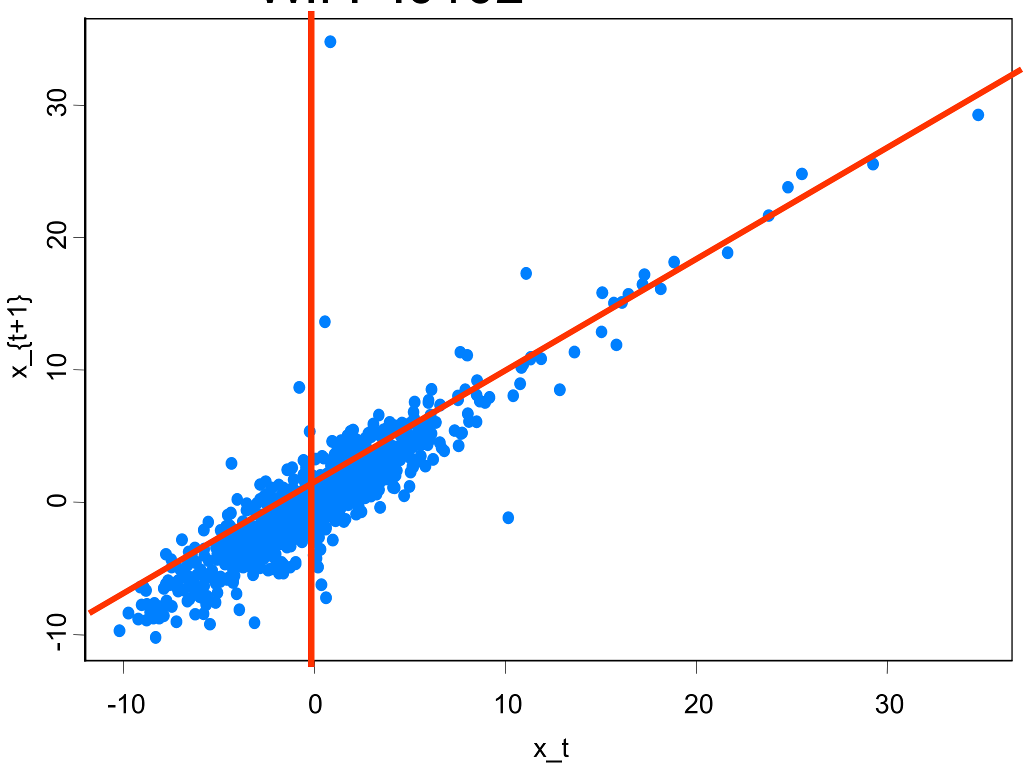
2. If $X_1 = X_2 > 0$, then $\mathbf{X} = (X_1, X_2)$ is multivariate regularly varying with index α and *spectral distribution*

$$P(\theta = (1/\sqrt{2}, 1/\sqrt{2})) = 1.$$

3. AR(1): $X_t = .9\, X_{t-1} + Z_t$, $\{Z_t\}$~IID symmetric stable (1.8)

$$\text{Distr of } \theta: \begin{cases} \pm(1,.9)/\text{sqrt}(1.81), & \text{W.P. } .9898 \\ \pm(0,1), & \text{W.P. } .0102 \end{cases}$$

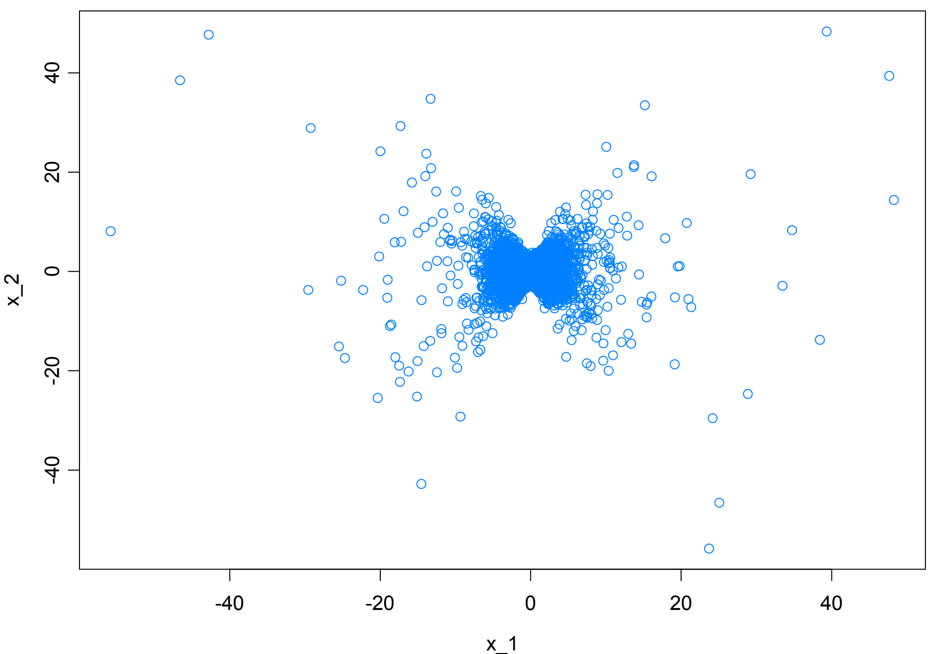
Figure: plot of (X_t, X_{t+1}) for realization of 10,000.

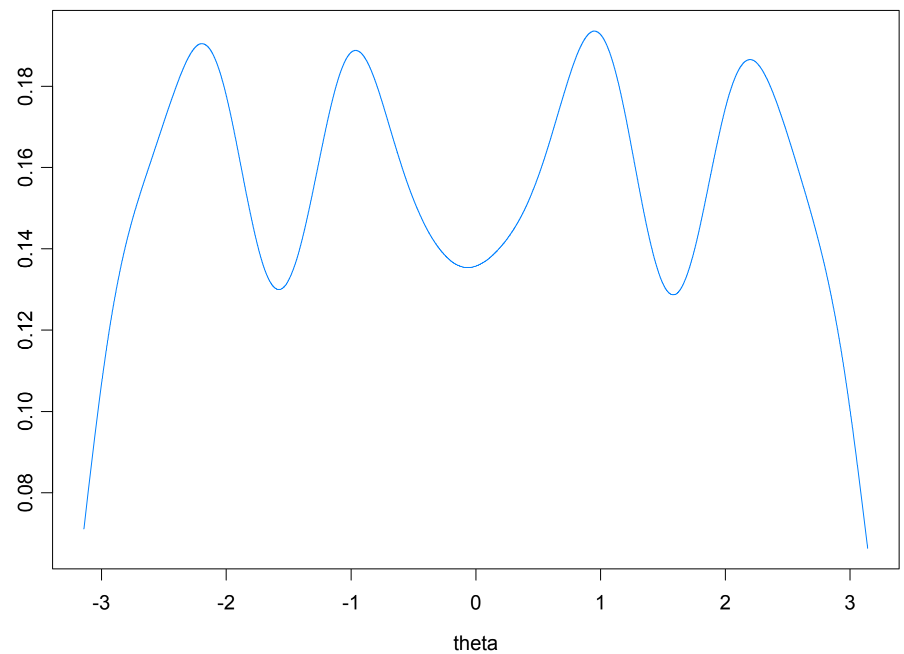


Examples (cont)

Example of ARCH(1): $\alpha_0=1$, $\alpha_1=1$, $\alpha=2$, $X_t=(\alpha_0+\alpha_1 X^2_{t-1})^{1/2}Z_t$, $\{Z_t\}$~IID

Figures: plots of (X_t, X_{t+1}) and estimated distribution of θ for realization of 10,000.





Examples (cont)

Example: SV model $X_t = \sigma_t Z_t$

Suppose $Z_t \sim RV(\alpha)$ and

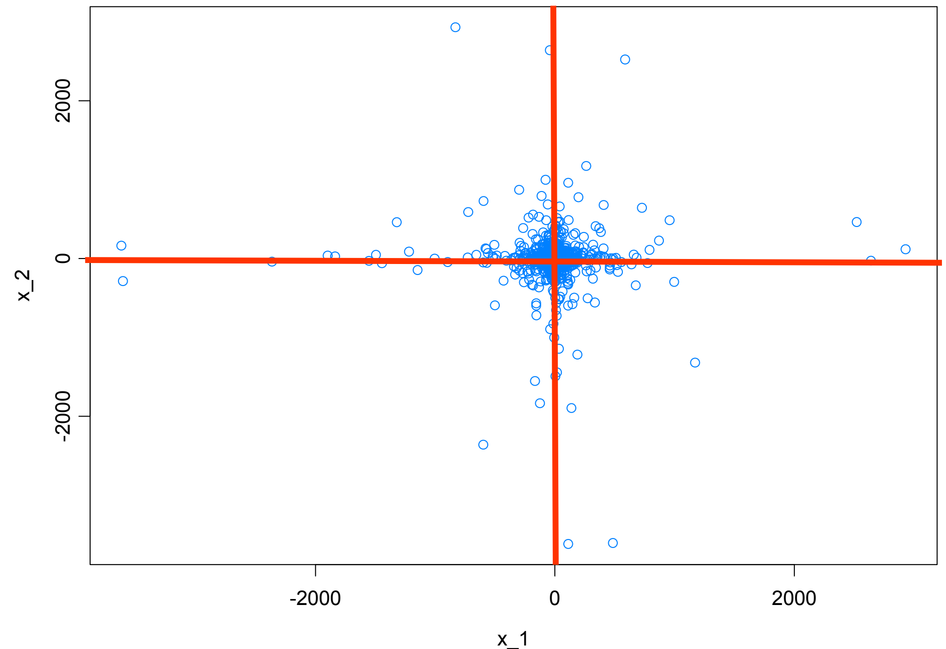
$$X_t = \sigma_t Z_t, \quad \log \sigma_t^2 = \phi_0 + \phi_1 \log \sigma_{t-1}^2 + \varepsilon_t, \quad \{\varepsilon_t\} \sim \text{IID N}(0,\sigma^2)$$

Then $\mathbf{Z}_n=(Z_1,\ldots,Z_n)'$ is regulary varying with index α and so is

$$\mathbf{X}_n = (X_1,\ldots,X_n)' = \text{diag}(\sigma_1,\ldots,\sigma_n)\,\mathbf{Z}_n$$

with spectral distribution concentrated on $(\pm 1,0)$, $(0, \pm 1)$.

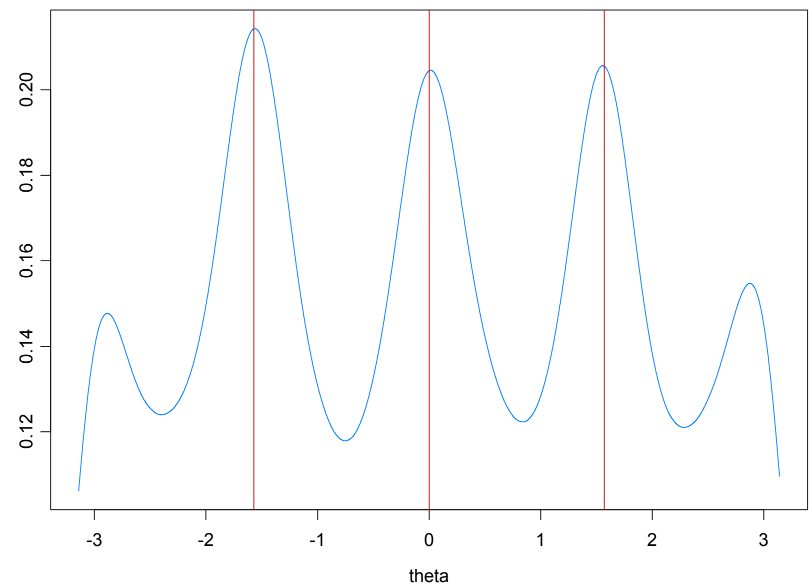
Figure: plot of (X_t,X_{t+1}) for realization of 10,000.

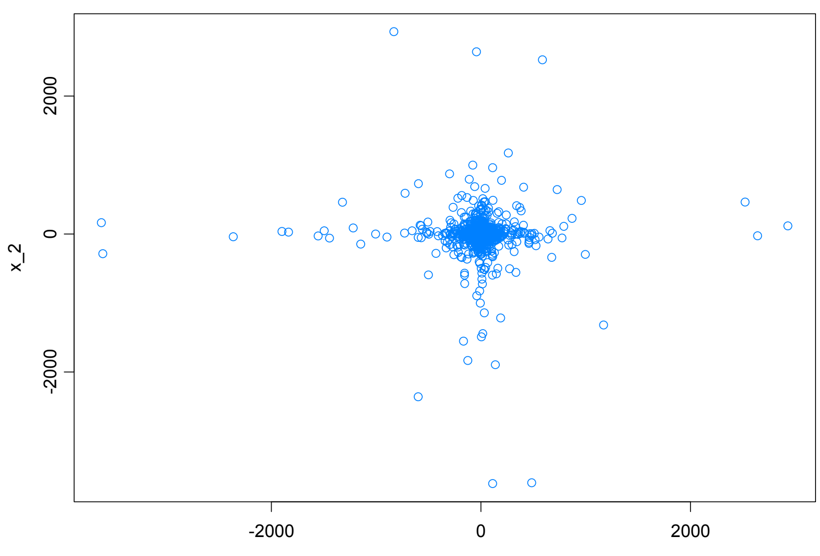


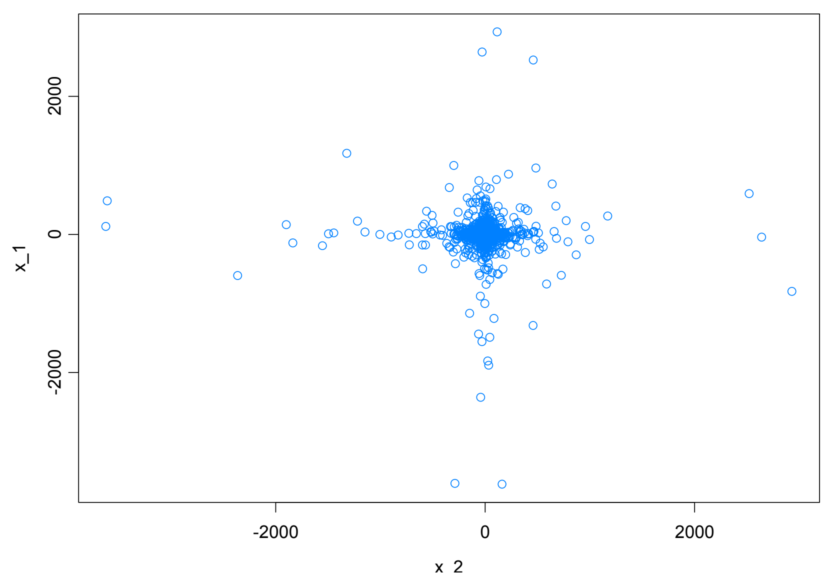
Examples (cont)

Example: SV model $X_t = \sigma_t Z_t$

- SV processes are time-reversible if log-volatility is Gaussian.
- Asymptotically time-reversible if log-volatility is nonGaussian







Regularly Varying Time Series

Definition: A stationary multivariate time series (X_t) is ***regularly varying*** with index α if all of its finite dimensional distributions are regularly varying with index α.

This means that for every $h > 0, \exists$ non-null measure μ_h such that μ_h is homogeneous, i.e., $\mu_h(t\,\cdot) = t^{-\alpha}\mu(\cdot)$ and

$$\frac{P(t^{-1}(X_1, \dots X_h) \in \cdot)}{P(|X_1| > t)} \stackrel{v}{\rightarrow} \mu_h(\cdot)$$

Note: 1) We use $P(|X_1| > t)$ in the denominator instead of $P(|X| > t)$, where $X = (X_1, \dots, X_h)$ as in the defn. But this does matter since,

$$\frac{P(|X| > t)}{P(|X_1| > t)} \rightarrow c_h := \mu_h(y: |y| > 1).$$

2) RV times series can be viewed as the heavy-tailed analogue of a Gaussian time series.

Regularly Varying Time Series

Examples:

1) Linear time series with RV-noise: $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, $(Z_t) \sim IID\ RV(\alpha)$

2) SRE $X_t = A_t X_{t-1} + B_t$ under the assumptions used earlier. Here we write

$$X_t = A_t A_{t-1} \cdots A_1 X_0 + R_t = \Pi_t X_0 + R_t$$

where $R_t = B_t + A_t B_{t-1} + \cdots + A_t A_{t-1} \cdots A_2 B_1$. Since $EA_1^{\alpha} = 1$ and the tail of R_t is much lighter than that of X_t, it follows from Breiman that

$$(X_0, \ldots, X_h) \sim X_0(\Pi_1, \ldots, \Pi_h)$$

is regularly varying. This implies GARCH time series are RV.

3) SV $X_t = \sigma_t Z_t$, $(Z_t) \sim IID\ RV(\alpha)$, and $\log \sigma_t = \sum_{j=0}^{\infty} \psi_j \eta_{t-j}$, $(\eta_t) \sim IID\ N(0, \sigma^2)$. In this, the measures μ_h are the same as those for an IID RV(α) sequence!

Point process convergence

Theorem (Davis & Hsing `95, Davis & Mikosch `97). Let $\{X_t\}$ be a stationary sequence of random *m*-vectors. Suppose

(i) finite dimensional distributions are jointly regularly varying (let $(\theta_{-k}, \ldots, \theta_k)$ be the vector in $S^{(2k+1)m-1}$ in the definition).

(ii) mixing condition $\mathcal{A}(a_n)$ or strong mixing.

(iii) $\lim_{k\to\infty}\limsup_{n\to\infty} P(\bigvee_{k\le|t|\le r_n} |\mathbf{X}_t| > a_n y \,\big|\, |\mathbf{X}_0| > a_n y) = 0.$

Then

$$\gamma = \lim_{k\to\infty} E(|\theta_0^{(k)}|^\alpha - \bigvee_{j=1}^{k} |\theta_j^{(k)}|)_+ / E|\theta_0^{(k)}|^\alpha$$

(extremal index)

exists. If $\gamma > 0$, then

$$N_n := \sum_{t=1}^{n} \varepsilon_{\mathbf{X}_t/a_n} \xrightarrow{d} N := \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \varepsilon_{\Gamma_i \mathbf{Q}_{ij}},$$

Point process convergence(cont)

- (Γ_i) are points of a Poisson process on $(0,\infty)$ with intensity function

$$\nu(dy)=\gamma\alpha y^{-\alpha-1}dy.$$

- $\sum_{j=1}^{\infty}\varepsilon_{Q_{ij}}$, $i \ge 1$, are iid point process with distribution Q, and Q is the weak limit of

$$\lim_{k\to\infty} E(|\theta_0^{(k)}|^\alpha - \bigvee_{j=1}^{k}|\theta_j^{(k)}|)_+ I_\bullet(\sum_{|t|\le k}\varepsilon_{\theta_t^{(k)}})/E(|\theta_0^{(k)}|^\alpha - \bigvee_{j=1}^{k}|\theta_j^{(k)}|)_+$$

Remarks:

1. GARCH and SV processes satisfy the conditions of the theorem.
2. Limit distribution for sample extremes and sample ACF follows from this theorem.

Extremes for GARCH and SV processes

Setup

- $X_t = \sigma_t Z_t$, $\{Z_t\} \sim$ IID (0,1)
- X_t is RV (α)
- Choose $\{b_n\}$ s.t. $nP(X_t > b_n) \to 1$

Then

$$P^n(b_n^{-1} X_1 \le x) \to \exp\{-x^{-\alpha}\}.$$

Then, with $M_n = \max\{X_1, \ldots, X_n\}$,

(i) GARCH:

$$P(b_n^{-1} M_n \le x) \to \exp\{-\gamma x^{-\alpha}\},$$

γ is extremal index ($0 < \gamma < 1$).

(ii) SV model:

$$P(b_n^{-1} M_n \le x) \to \exp\{-x^{-\alpha}\},$$

extremal index $\gamma = 1$ no clustering.

Extremes for GARCH and SV processes (cont)

(i) GARCH: $P(b_n^{-1}M_n \le x) \to \exp\{-\gamma x^{-\alpha}\}$

(ii) SV model: $P(b_n^{-1}M_n \le x) \to \exp\{-x^{-\alpha}\}$

Remarks about extremal index.

(i) $\gamma < 1$ implies clustering of exceedances

(ii) Numerical example. Suppose c is a threshold such that

$$P^n(b_n^{-1}X_1 \le c) \sim .95$$

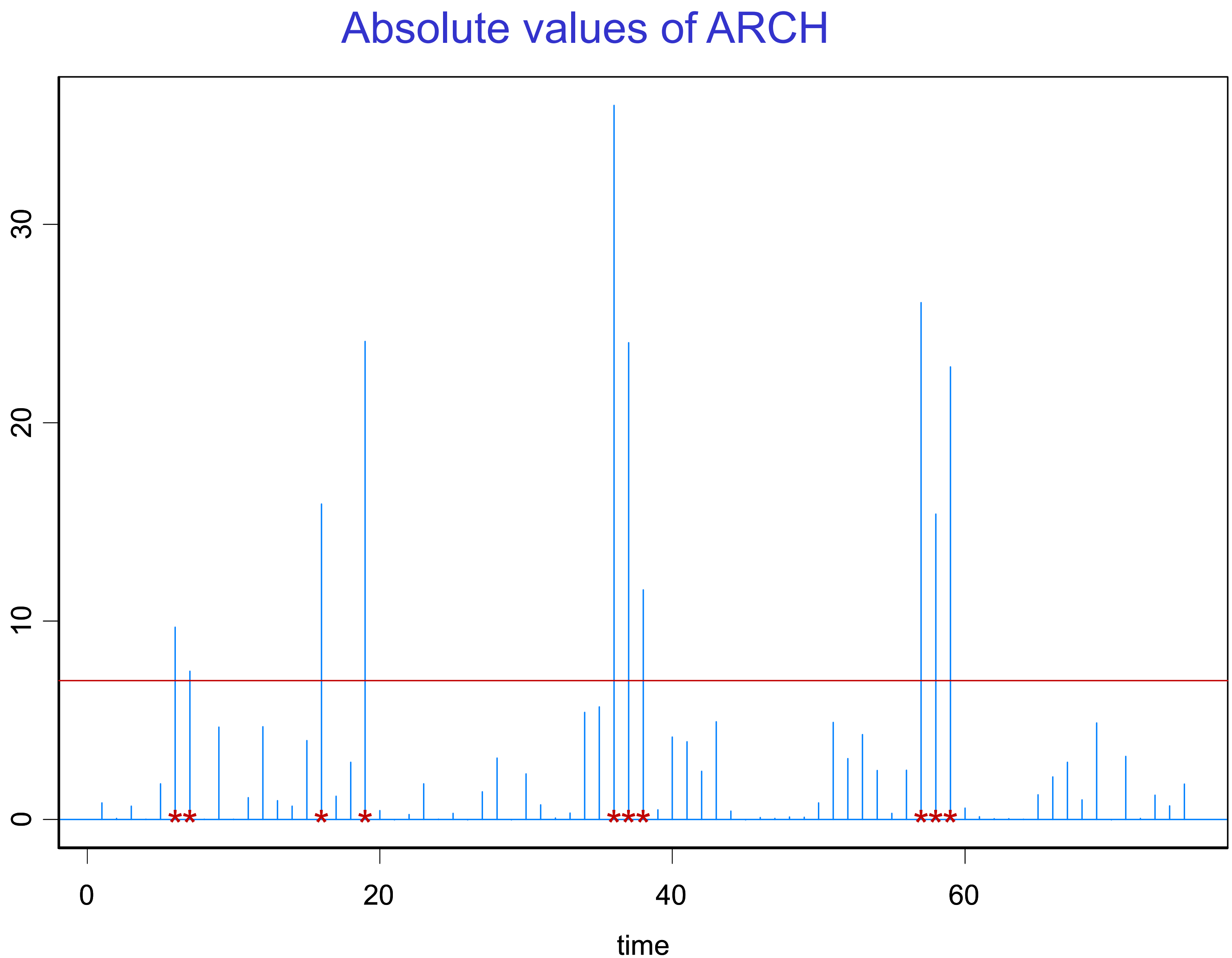
Then, if $\gamma = .5$, $P(b_n^{-1}M_n \le c) \sim (.95)^{.5} = .975$

(iii) $1/\gamma$ is the *mean cluster size* of exceedances.

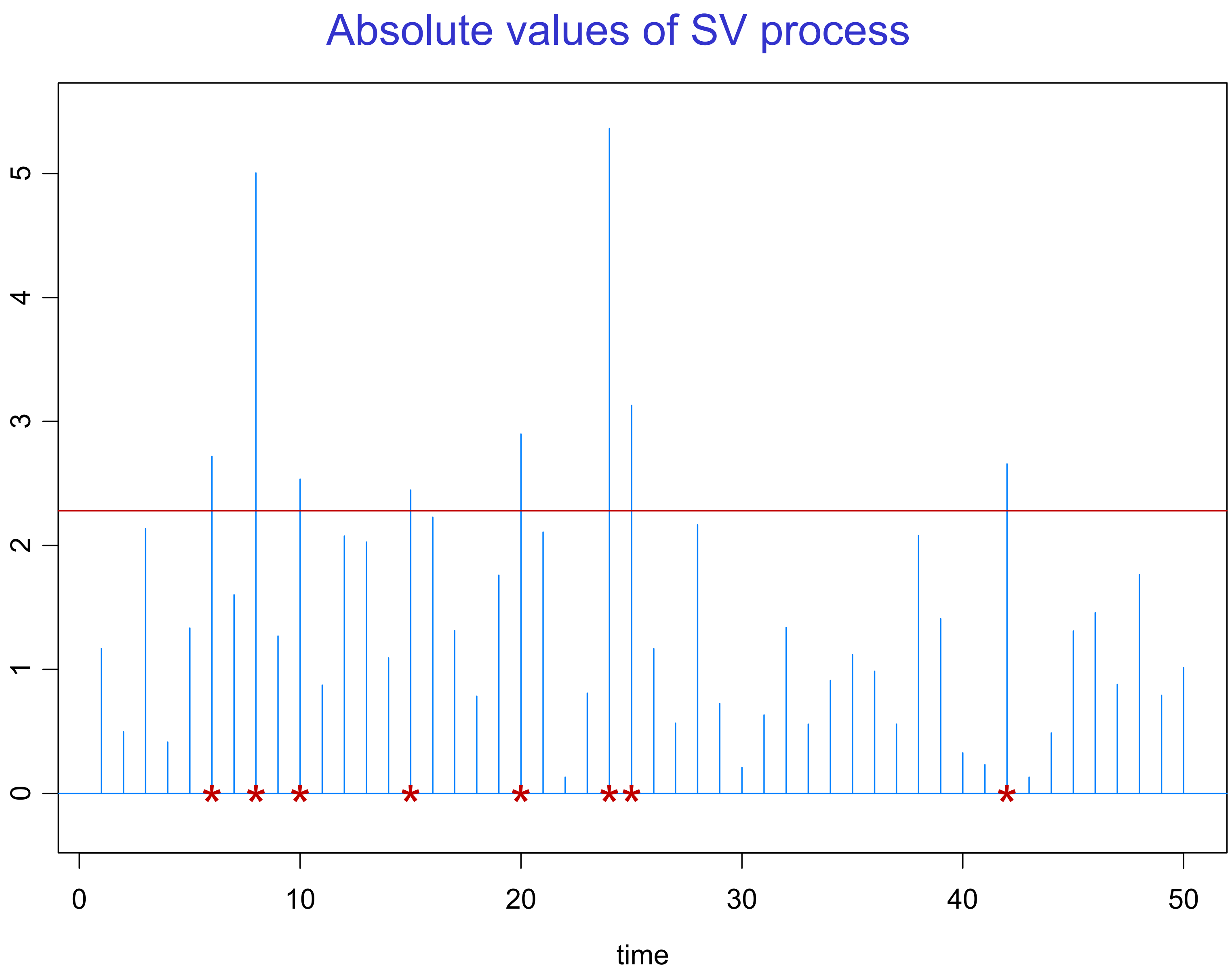
(iv) Use γ to *discriminate* between GARCH and SV models.

(v) Even for the light-tailed SV model (i.e., $\{Z_t\}$ ~IID N(0,1), the *extremal index* is 1 (see Breidt and Davis `98)

Extremes for GARCH and SV processes (cont)



Extremes for GARCH and SV processes (cont)



Summary of results for ACF of GARCH(p,q) and SV models

GARCH(p,q)

$\alpha \in (0,2)$:

$$(\hat{\rho}_X(h))_{h=1,\ldots,m} \xrightarrow{d} (V_h / V_0)_{h=1,\ldots,m},$$

$\alpha \in (2,4)$:

$$\left(n^{1-2/\alpha}\hat{\rho}_X(h)\right)_{h=1,\ldots,m} \xrightarrow{d} \gamma_X^{-1}(0)\left(V_h\right)_{h=1,\ldots,m}.$$

$\alpha \in (4,\infty)$:

$$\left(n^{1/2}\hat{\rho}_X(h)\right)_{h=1,\ldots,m} \xrightarrow{d} \gamma_X^{-1}(0)\left(G_h\right)_{h=1,\ldots,m}.$$

Remark: Similar results hold for the sample ACF based on $|X_t|$ and X_t^2.

Summary of results for ACF of GARCH(p,q) and SV models (cont)

SV Model

$\alpha \in (0,2)$:

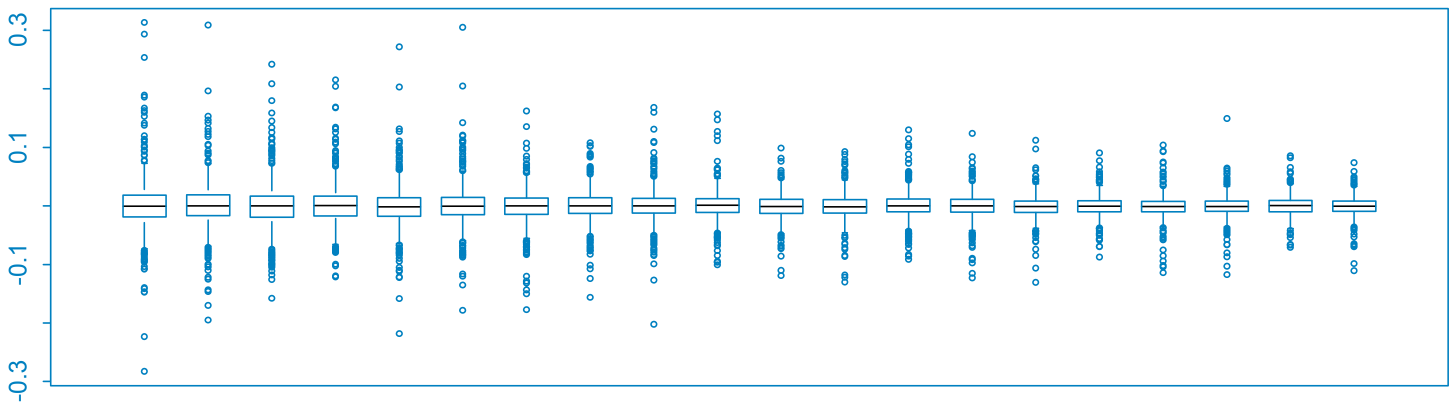
$$\left(n/\ln n\right)^{1/\alpha}\hat{\rho}_X(h) \xrightarrow{d} \frac{\|\sigma_1\sigma_{h+1}\|_\alpha}{\|\sigma_1\|_\alpha^2}\frac{S_h}{S_0}.$$

$\alpha \in (2, \infty)$:

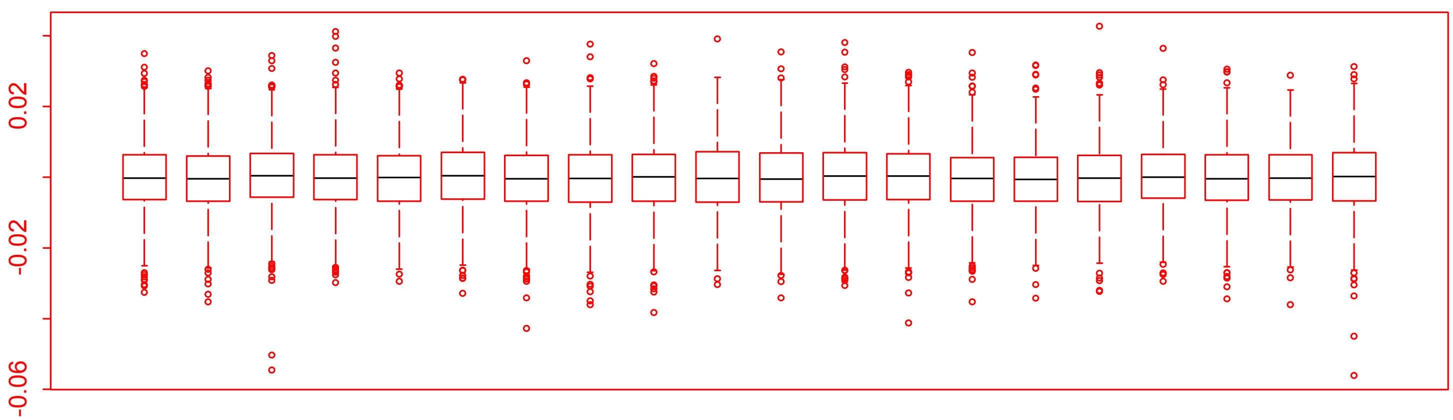
$$\left(n^{1/2}\hat{\rho}_X(h)\right)_{h=1,\ldots,m} \xrightarrow{d} \gamma_X^{-1}(0)\left(G_h\right)_{h=1,\ldots,m}.$$

Sample ACF for GARCH and SV Models (1000 reps)

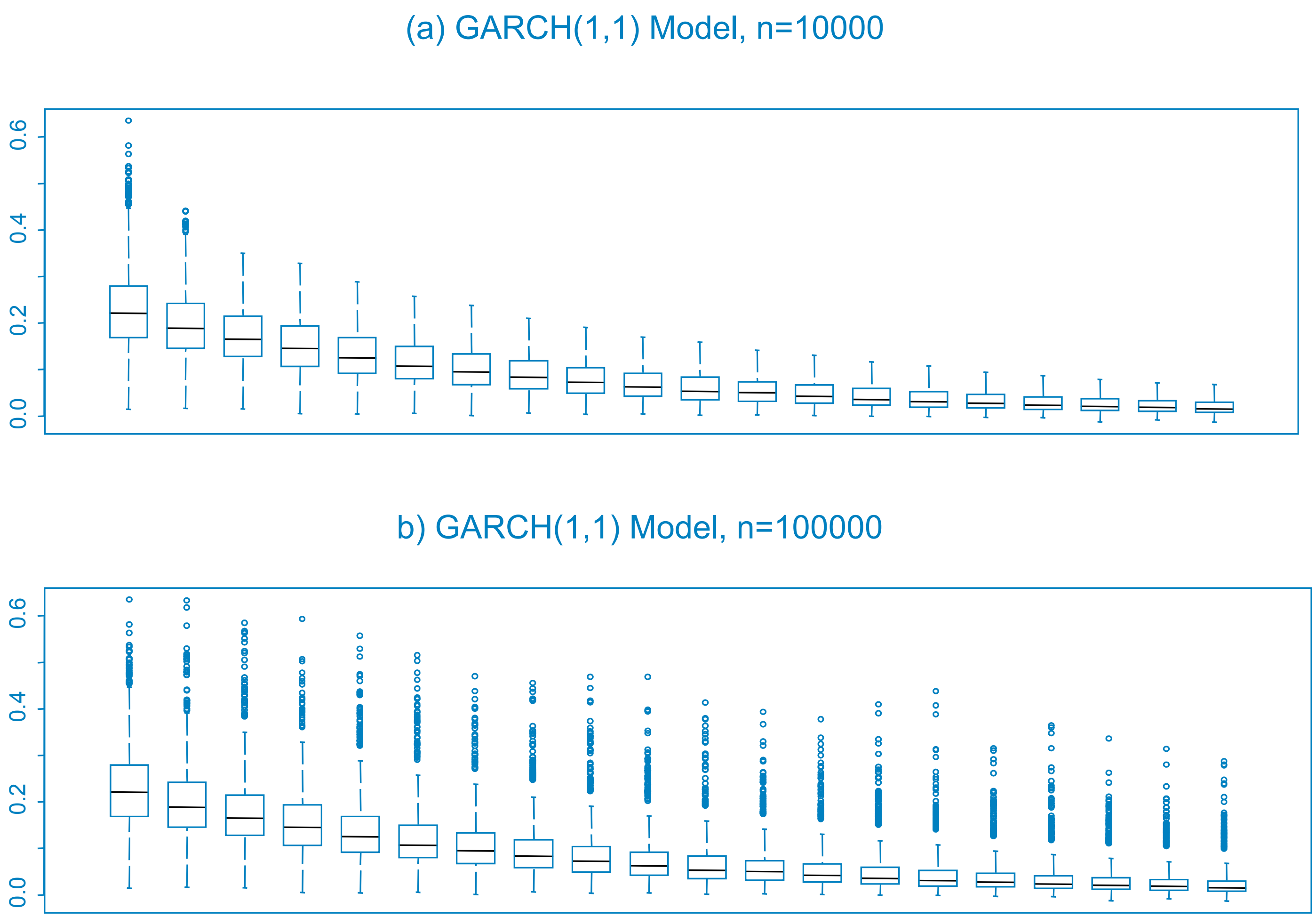
(a) GARCH(1,1) Model, n=10000



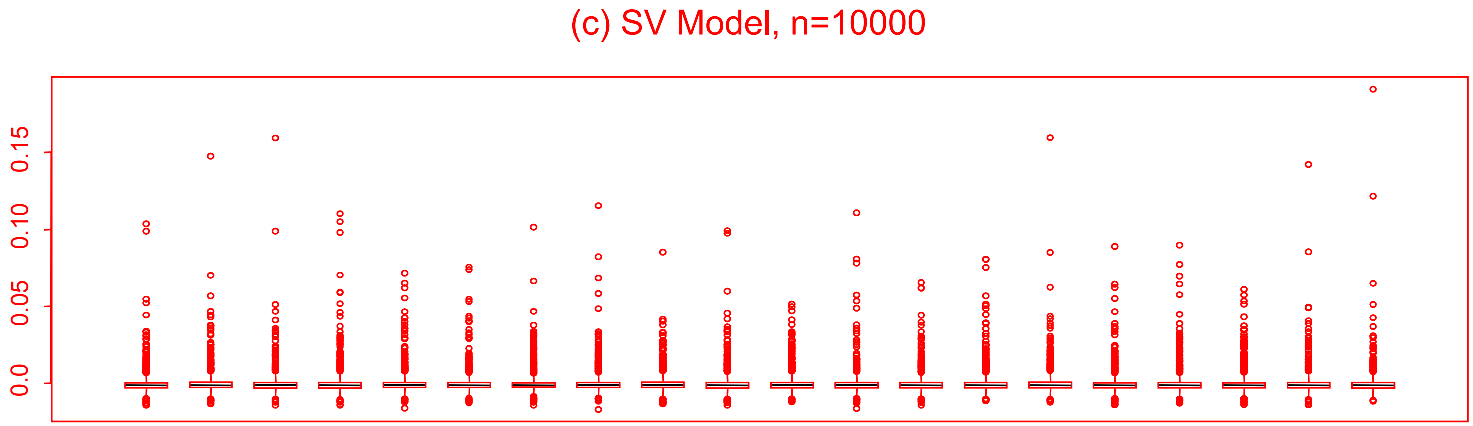
(b) SV Model, n=10000

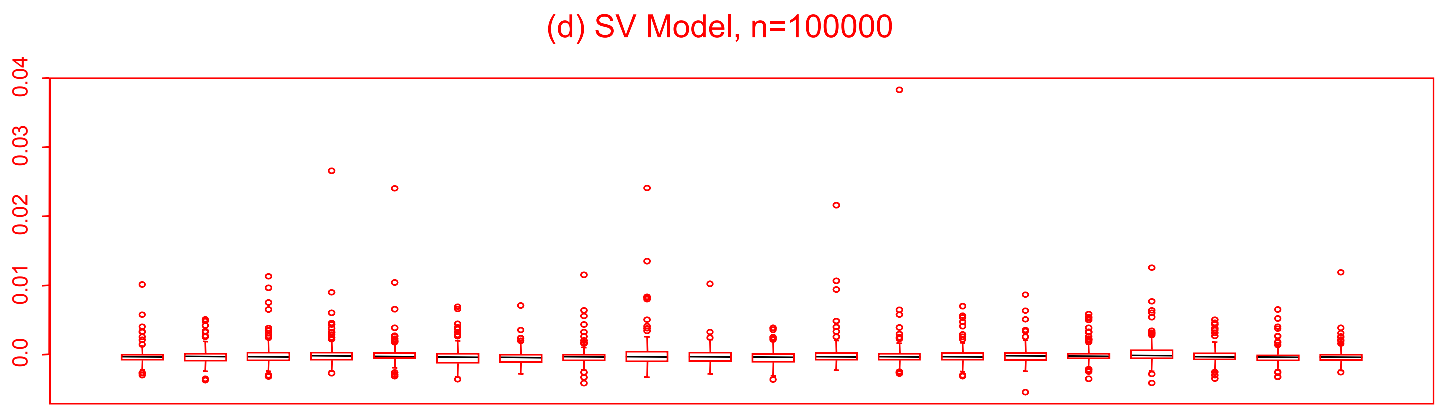


Sample ACF for Squares of GARCH (1000 reps)

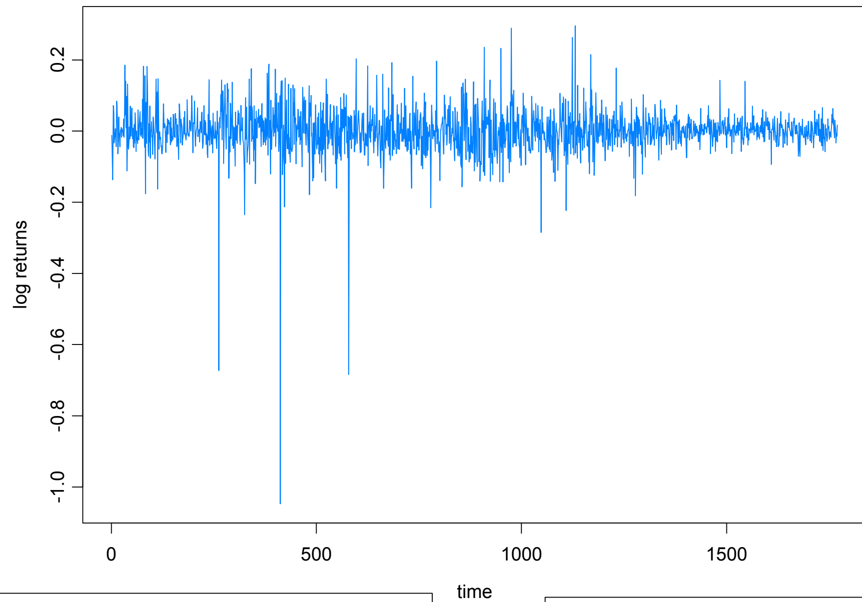


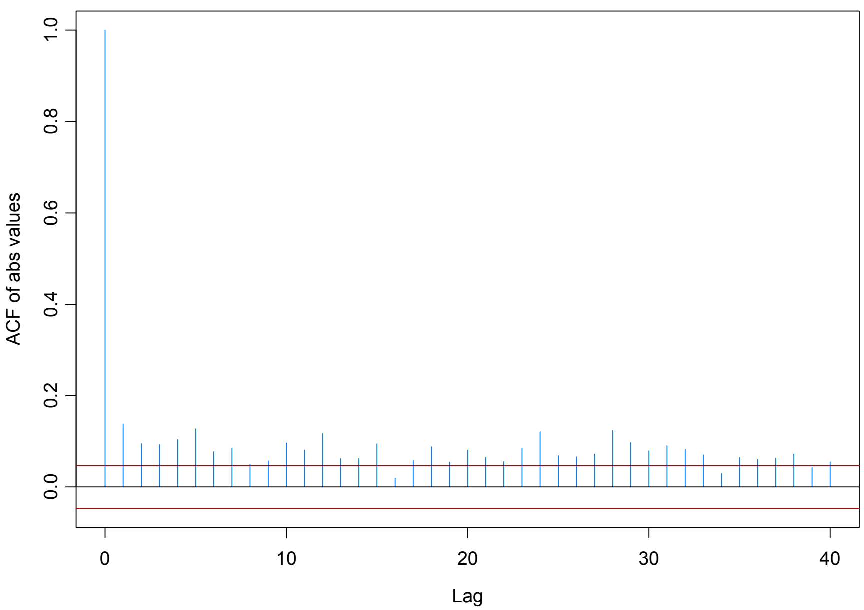
Sample ACF for Squares of SV (1000 reps)

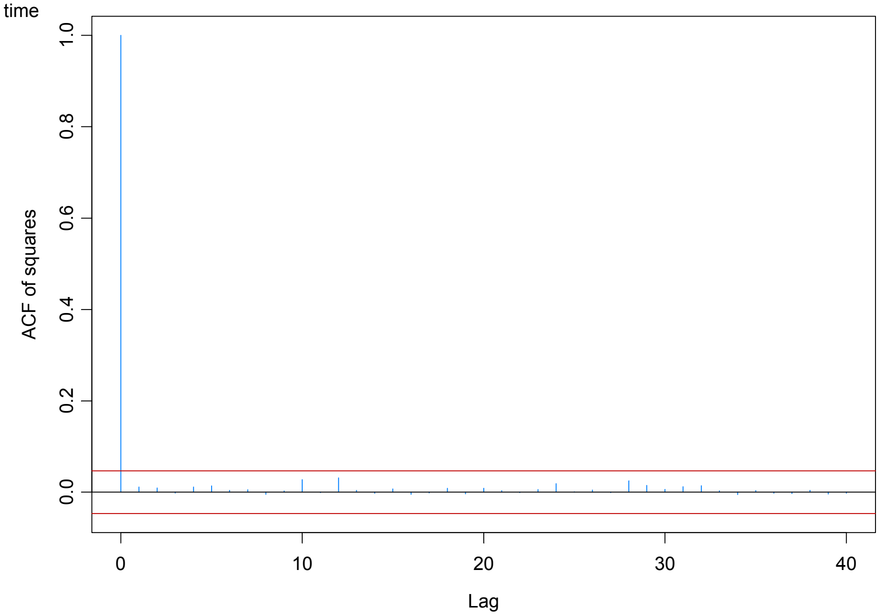




Example: Amazon-returns (May 16, 1997 – June 16, 2004)





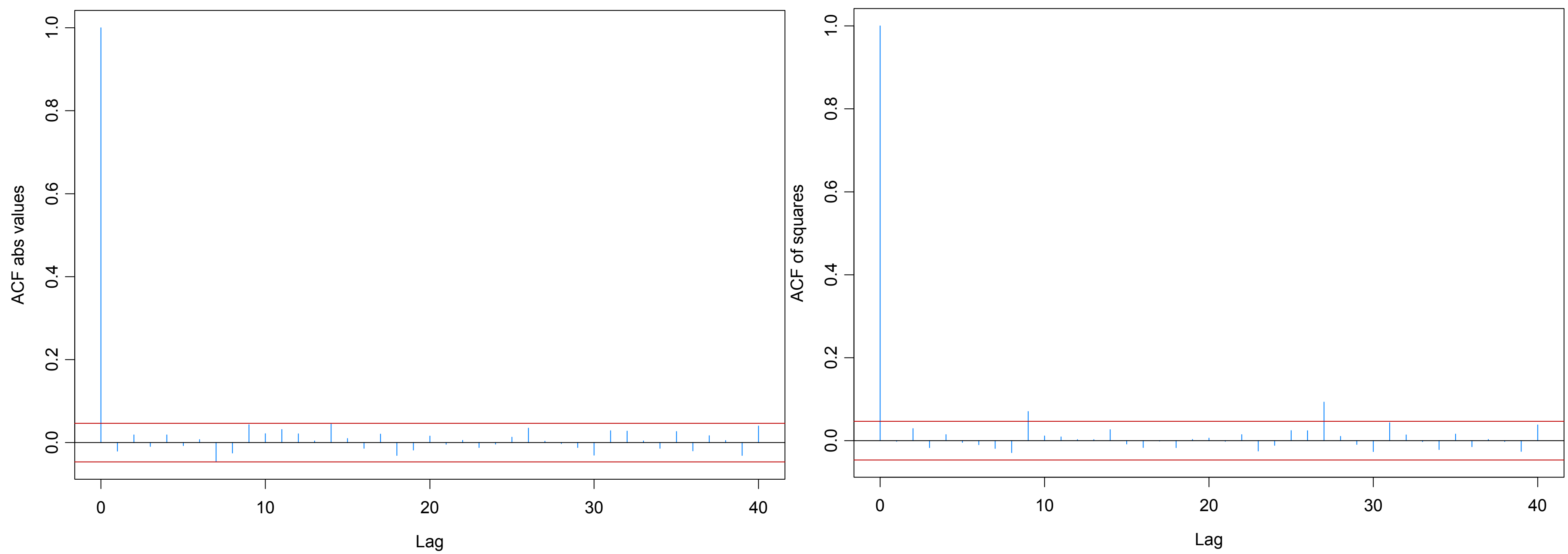


Amazon returns (GARCH model)

GARCH(1,1) model fit to Amazon returns:

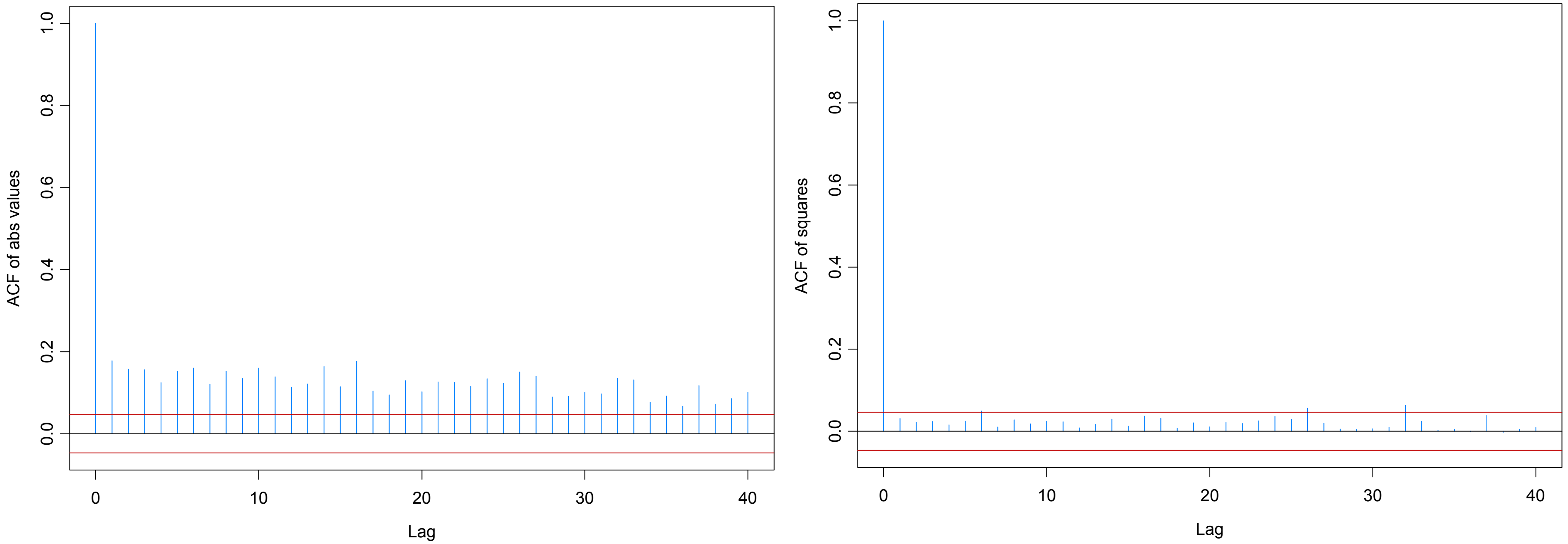
α_0= .00002493, α_1= .0385, β_1 = .957, $X_t=(\alpha_0+\alpha_1 X^2_{t-1})^{1/2}Z_t$, $\{Z_t\}$~IID t(3.672)

Simulation from GARCH(1,1) model



Amazon returns (SV model)

Stochastic volatility model fit to Amazon returns:



The Extremogram

The extremogram of a stationary time series $\{X_t\}$ can be viewed as the analogue of the correlogram in time series for measuring dependence in extremes (see Davis and Mikosch (2009)).

Definition: For two sets A & B ***bounded away from 0***, the ***extremogram*** is defined as

$$\rho_{A,B}(h) = \lim_{x\to\infty} P(\mathbf{X}_h \in xB \mid \mathbf{X}_0 \in xA)$$

$$= \lim_{x\to\infty} P(\mathbf{X}_0 \in xA, \mathbf{X}_h \in xB)/P(\mathbf{X}_0 \in xA),$$

for $h = 0, 1, \ldots,$ provided the limit exists, where $\mathbf{X}_h=(X_h,X_{h+1},\ldots,X_{h+k})$.

Remark: This definition requires that the limit exists.

a) exists for heavy-tailed time series (see forthcoming slide)

b) exists for some light-tailed time series w/ special choices of A and B.

c) extremal dependence ***depends*** on the choice of sets A & B.

The Extremogram (cont)

If one takes A=B=$(1,\infty)$ and $k = 0$, then

$$\rho_{A,B}(h) = \lim_{x\to\infty} P(X_h > x \mid X_0 > x) = \lambda(X_0, X_h)$$

often called the ***extremal dependence coefficient*** ($\lambda = 0$ means independence or asymptotic independence).

Other choices of A and B can lead to interesting extremograms:

- $P(X_1 < -x \mid X_0 < -x)$ (negative return followed by a neg return)
- $P(X_1 > x \mid X_0 < -x)$ (neg return followed by a pos return)
- $P(X_1 + \cdots + X_4 > 2x \mid X_0 < -x)$ (neg return followed by a big pos return aggregated over next 4 days)
- $P(X_1 > x, \ldots, X_4 > x \mid X_0 > x)$ (pos return followed by a pos return in next 4 days)
- $P(\min\{X_2, X_3, X_4\} > x \mid X_0 > x, X_1 > x)$ (2 pos returns$\Rightarrow$pos return)

The Extremogram: examples

Let A = B = $(1,\infty)$, then

$$\rho_{A,B}(h) = \lim_{x\to\infty} P(X_0 > x, X_h > x)/P(X_0 > x)$$

Gaussian Processes: In this case,

$$\rho_{A,B}(h) = 0 \text{ for all } h > 0 \quad \text{(asymptotic independence).}$$

GARCH: In this case

$$\rho_{A,B}(h) > 0 \text{ for all } h > 0,$$

but decays to 0 geometrically fast.

SV process: $X_t = \sigma_t Z_t, \quad \log \sigma_t^2 = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \; \{\varepsilon_t\} \sim \text{IIDN}(0,\sigma^2)$

In this case,

$$\rho_{A,B}(h) = 0 \text{ for all } h > 0.$$

The Extremogram: examples

Let $A = B = (1,\infty)$, then

$$\rho_{A,B}(h) = \lim_{x\to\infty} P(X_0 > x, X_h > x)/P(X_0 > x)$$

AR(1): $X_t = \phi X_{t-1} + Z_t$, $\{Z_t\}$~IID Cauchy. Then

$$\rho_{A,B}(h) = \max(0, \phi^h).$$

Note if $\phi < 0$, then extremogram alternates between positive #'s and 0

MaxMA(2): Let (Z_t) be iid with Pareto distribution, i.e., $P(Z_1 > x) = x^{-\alpha}$ for $x \geq 1$, and set $X_t = \max(Z_t, Z_{t-1}, Z_{t-2})$. Then

$$\begin{aligned} \rho_{A,B}(h) &= 1 \quad \text{for } h = 0. \\ &= 2/3 \quad \text{for } h = 1 \\ &= 1/3 \quad \text{for } h = 2 \\ &= 0, \quad \text{for } h > 2. \end{aligned}$$

Regular Variation and the Extremogram

Facts

1. The extremogram of a RV stationary time series $\{X_t\}$ exists for all choices of sets A & B bounded away from the origin.

2. Many heavy-tailed time series (GARCH and SV) are regularly varying.

The Empirical Extremogram

A natural estimator of the extremogram,

$$\rho_{A,B}(h) = \lim_{x\to\infty} P(X_h \in xB \mid X_0 \in xA)$$

based on observations, $X_1, \ldots, X_n$, is the empirical extremogram defined by

$$\hat{\rho}_{A,B}(h) = \frac{\frac{m}{n}\sum_{t=1}^{n-h} I_{\{a_m^{-1}X_t \in A,\, a_m^{-1}X_{t+h} \in B\}}}{\frac{m}{n}\sum_{t=1}^{n} I_{\{a_m^{-1}X_t \in A\}}},$$

where a_m is the $1 - 1/m$ quantile of $|X_t|$. For the theory to work, need

$$m_n \to \infty \text{ with } m/n \to 0.$$

Under suitable mixing conditions, a CLT for the empirical estimate is established in D&M (2009).

The Empirical Extremogram — central limit theorem

$$\hat{\rho}_{A,B}(h) = \frac{\frac{m}{n}\sum_{t=1}^{n-h} I_{\{a_m^{-1}X_t \in A,\, a_m^{-1}X_{t+h} \in B\}}}{\frac{m}{n}\sum_{t=1}^{n} I_{\{a_m^{-1}X_t \in A\}}}$$

After first establishing a joint CLT for the numerator and denominator, we obtain the limit result

$$(n/m)^{1/2}(\hat{\rho}_{A,B}(h) - \rho_m(h)) \to_d N(0, \sigma^2(A,B)),$$

where $\rho_m(h)$ is the ratio of expectations (***pre-asymptotic extremogram***),

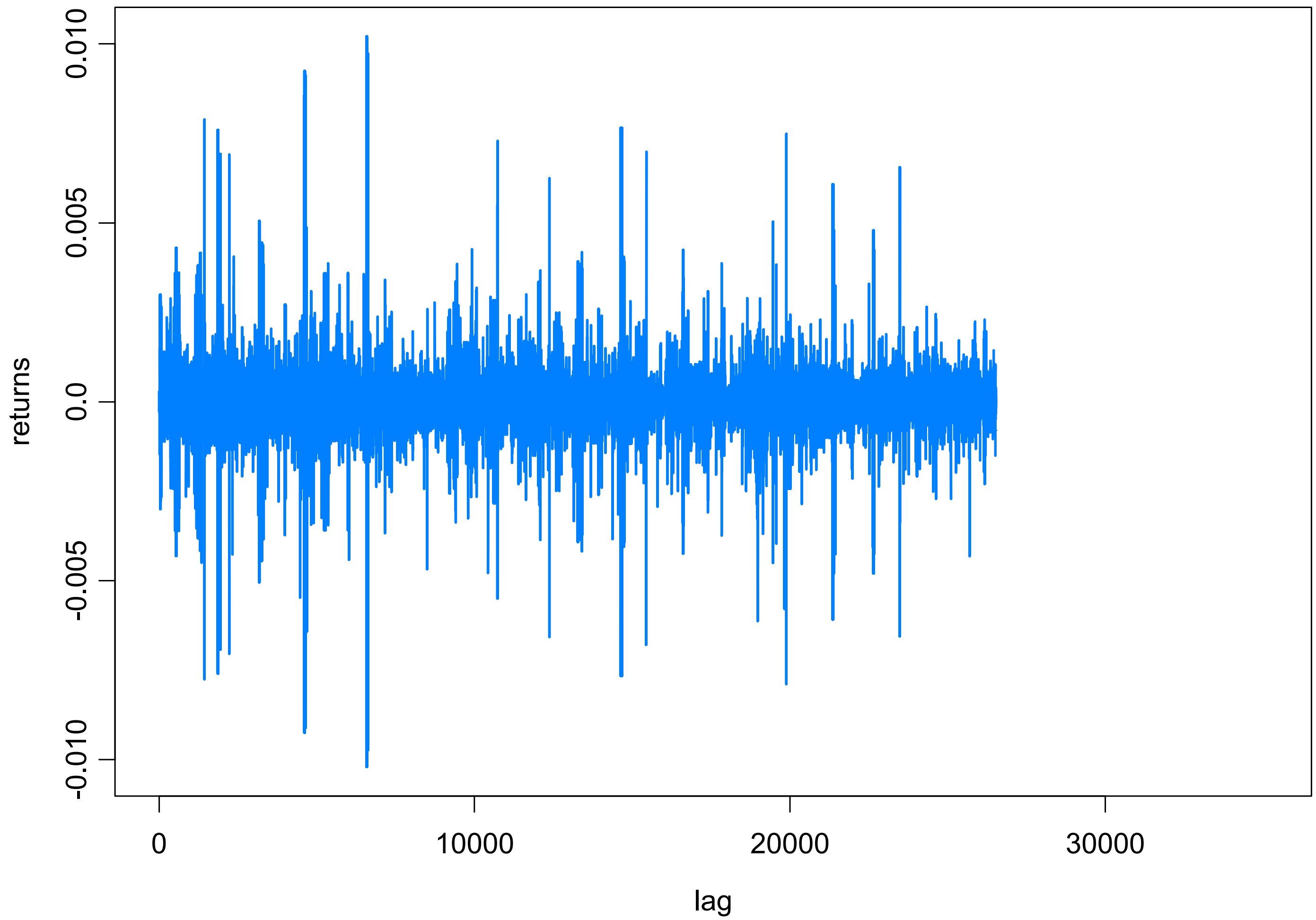
$$P\,(a_m^{-1}X_0 \in A,\ a_m^{-1}X_h \in B) / P\,(a_m^{-1}X_0 \in A).$$

Now provided a bias condition, such as

$$(n/m)^{1/2}\left(mP\,(a_m^{-1}X_0 \in A,\ a_m^{-1}X_h \in B) - \mu_h(A \times B)\right) \to 0,$$

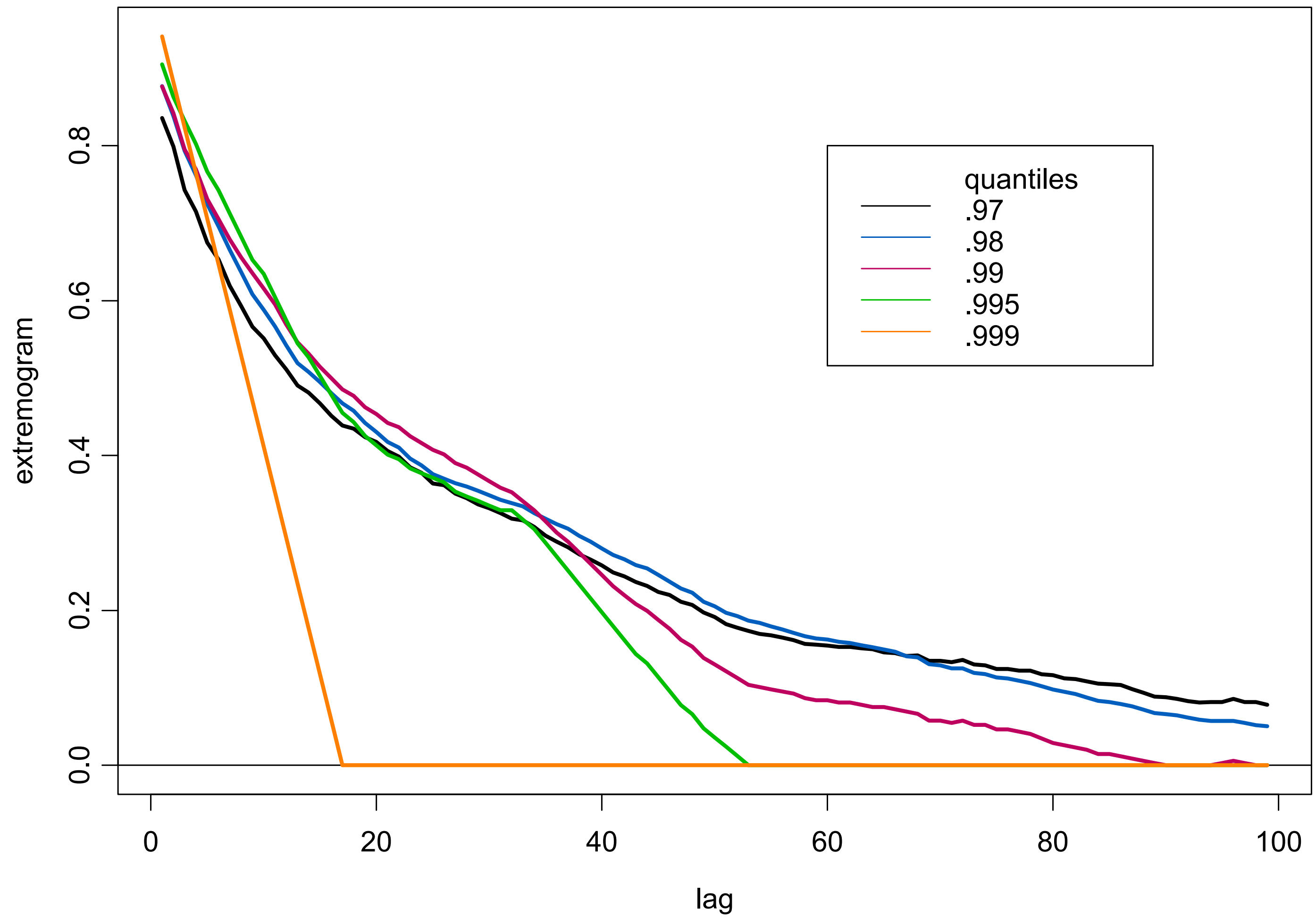
holds, then $\rho_m(h)$ can be replaced with $\rho_{A,B}(h)$. This condition can often be difficult to check.

Application to Five-Minute Return Data (US/DM) exchange



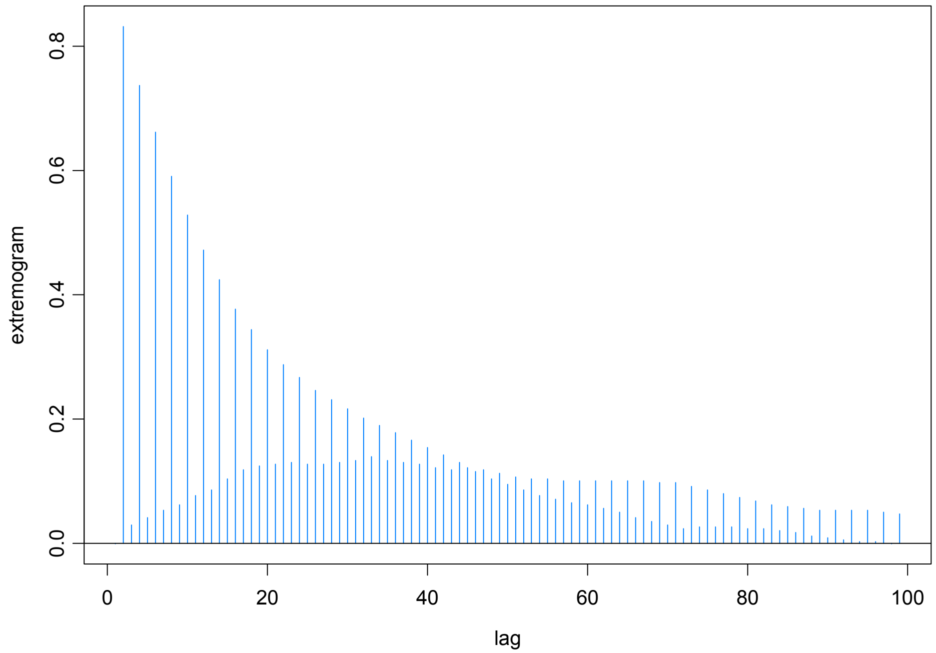
Application to Five-Minute Return Data (US/DM) exchange

Extremogram absolute values: choice of threshold a_m



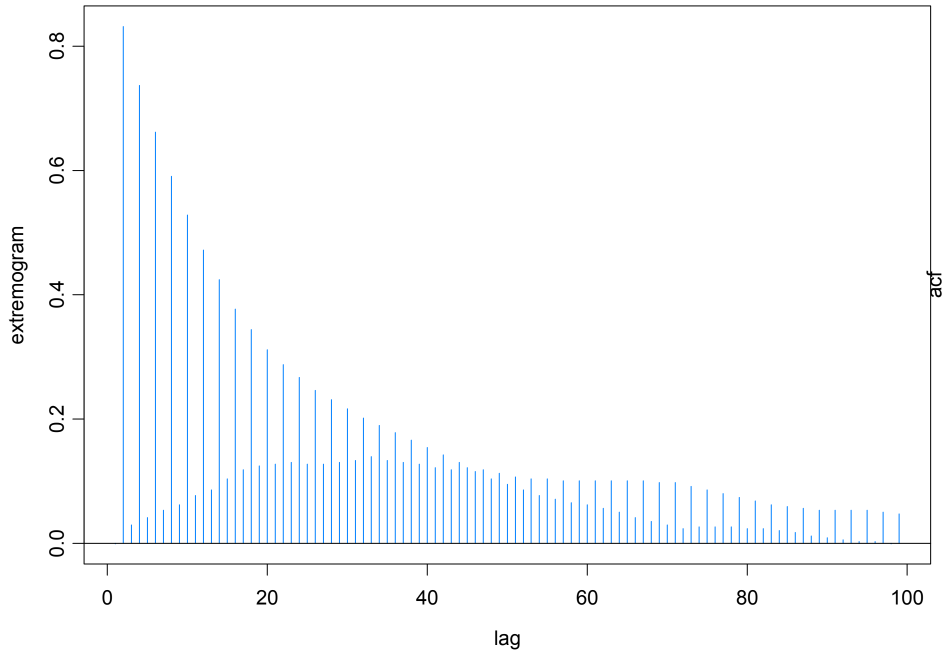
Application to Five-Minute Return Data (US/DM) exchange

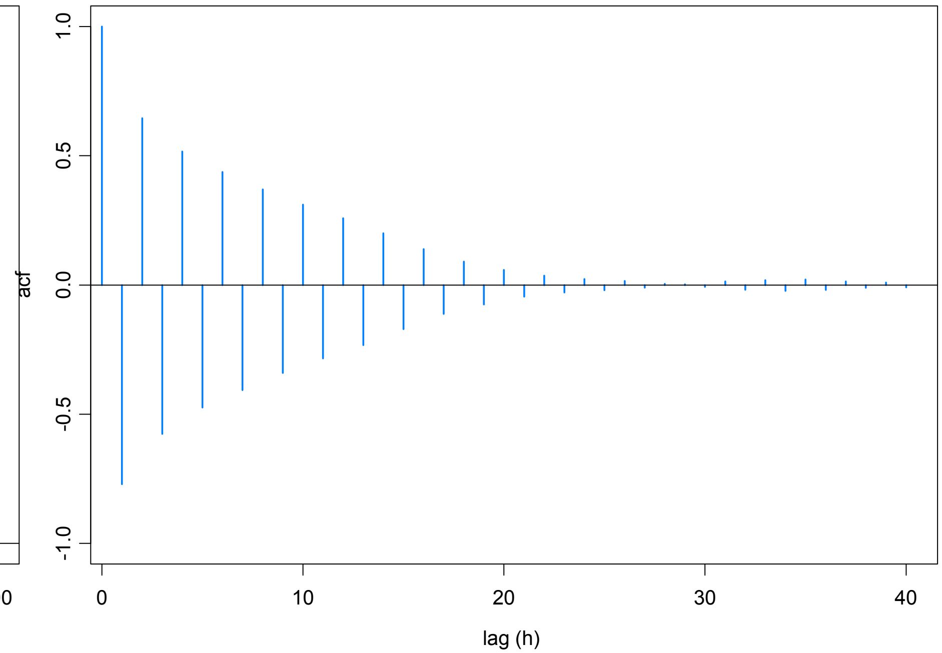
Extremogram A=B=$(1,\infty)$



Application to Five-Minute Return Data (US/DM) exchange

Extremogram A=B=$(1,\infty)$

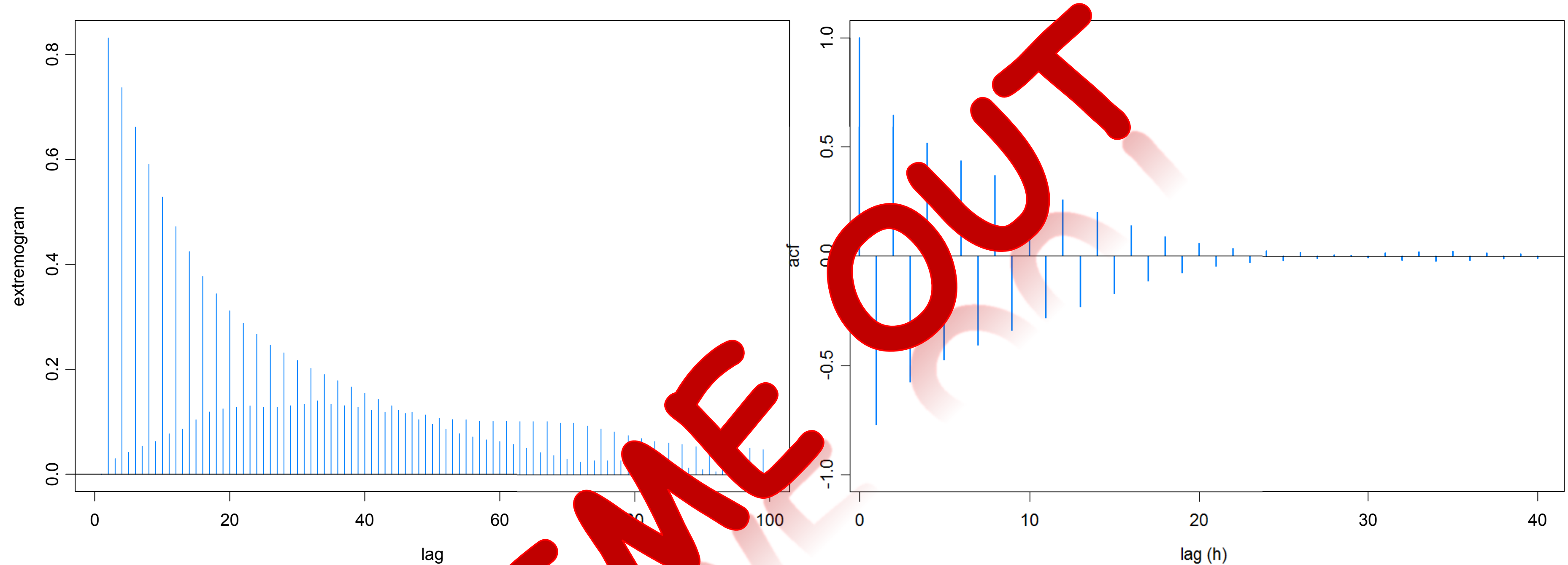




Best fitting AR model is of order 18; refine with nonzero coefficients at lags 1, 2, 3, 5, 6, 7, 11, 13, 14, 16, and18.

Application to Five-Minute Return Data (US/DM) exchange

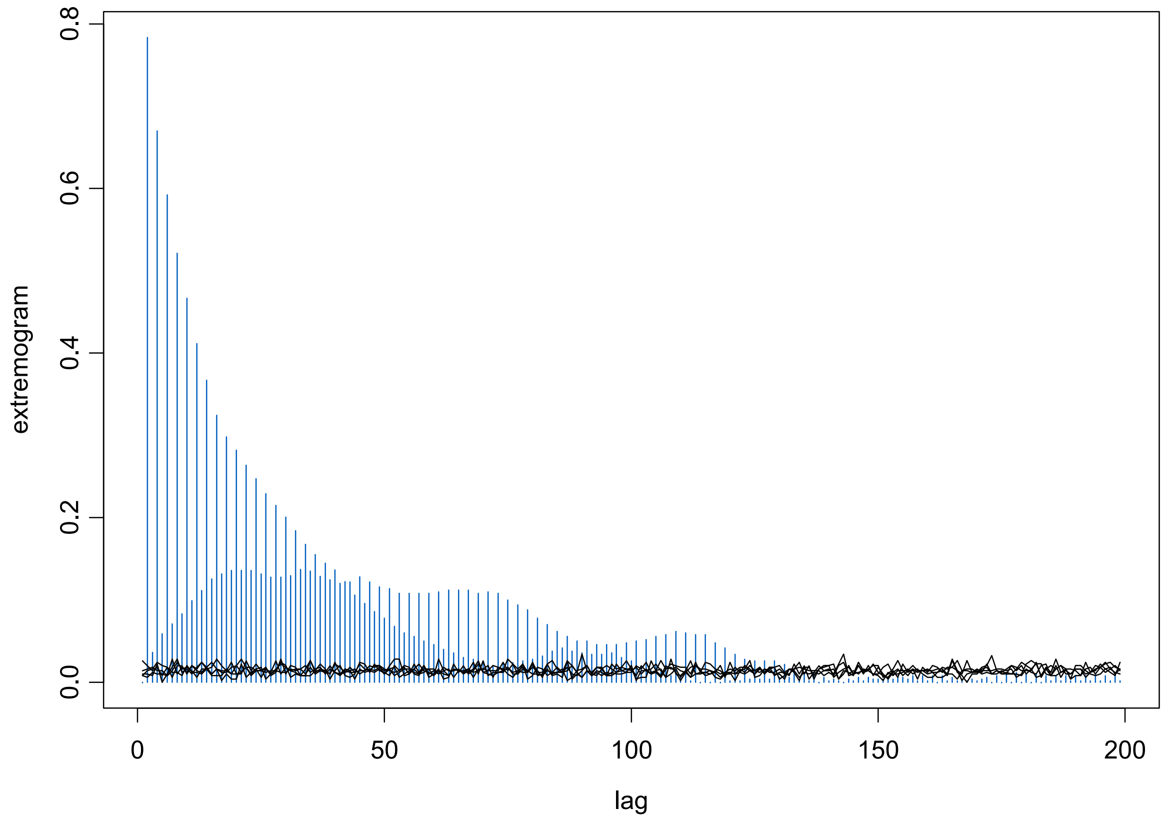
Extremogram $A=B=(1,\infty)$



Best fitting AR model is of order 18; refine with nonzero coefficients at lags 1, 2, 3, 5, 6, 7, 11, 13, 14, 16, and18.

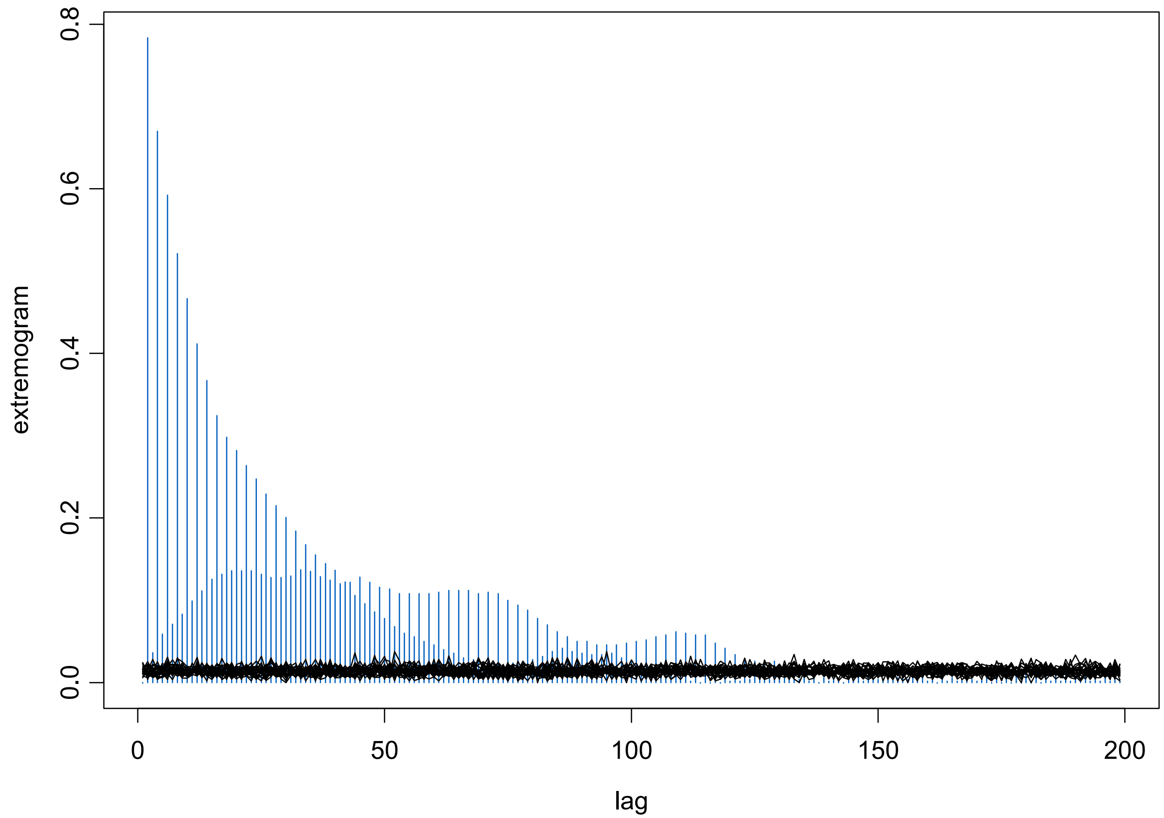
Time out: Resampling and Testing for Serial Dependence

A natural way (*not often used in time series*) for testing serial correlation is to compute the ACF for random permutations of the data. If the sample ACF appears more ***extreme*** than the ACFs based on random permutations, then there is evidence of serial correlation. We apply the same idea to the extremogram.

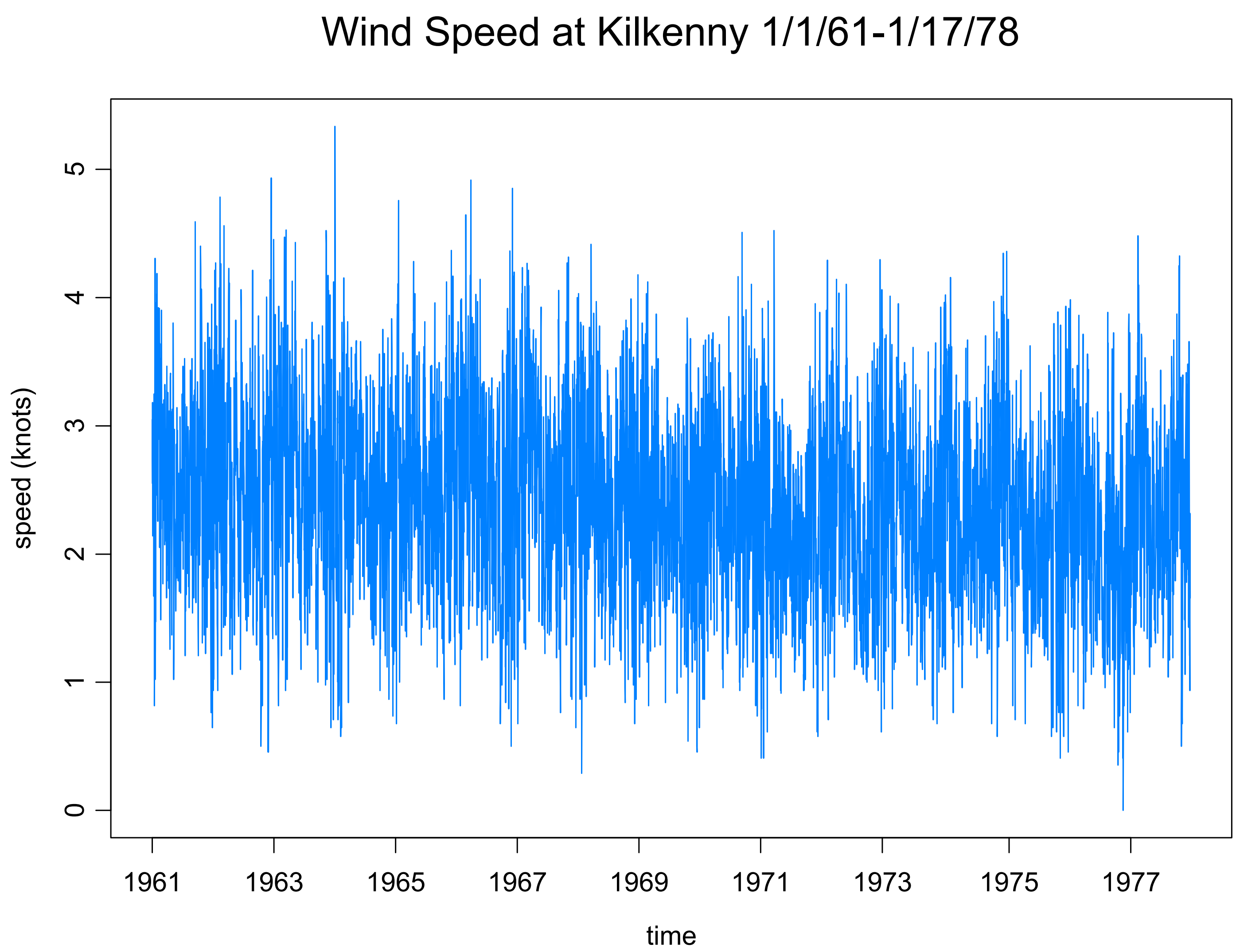


Time out: Resampling and Testing for Serial Dependence

A natural way (*not often used in time series*) for testing serial correlation is to compute the ACF for random permutations of the data. If the sample ACF appears more ***extreme*** than the ACFs based on random permutations, then there is evidence of serial correlation. We apply the same idea to the extremogram.

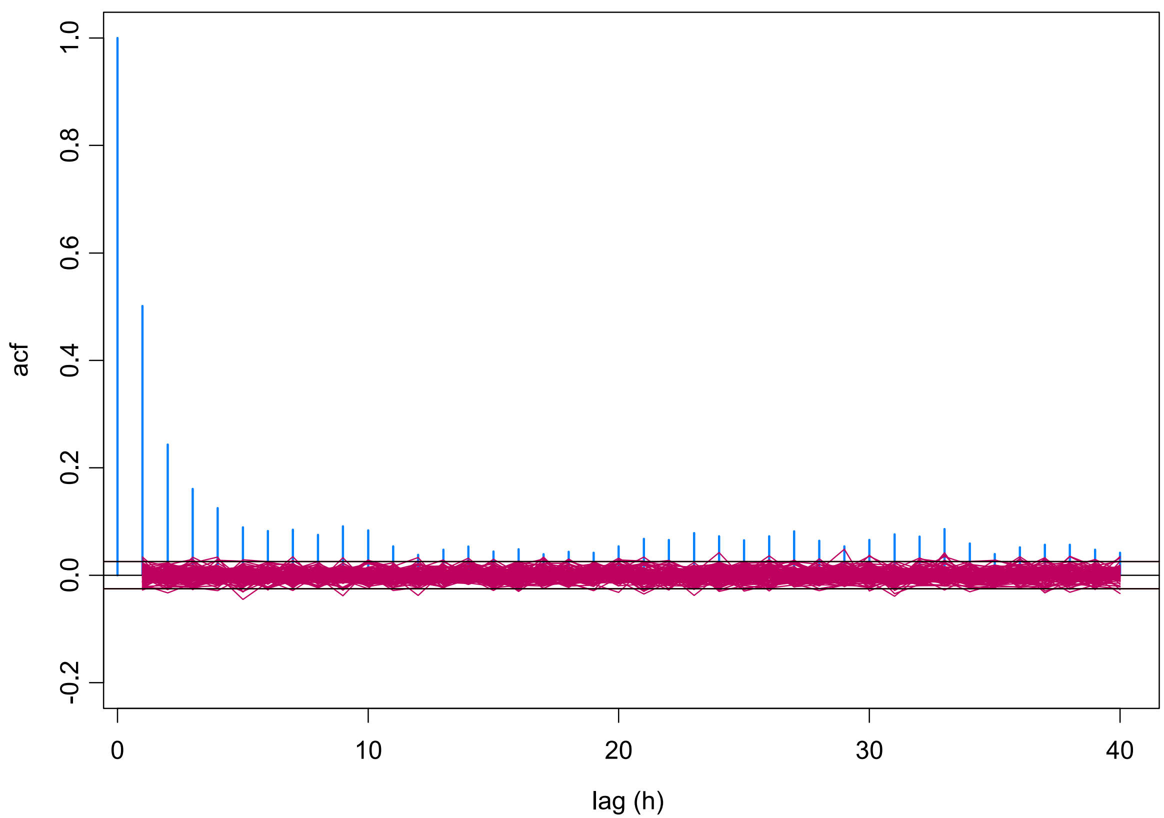


Time out: Illustration with ACF (Windspeed at Kilkenny)

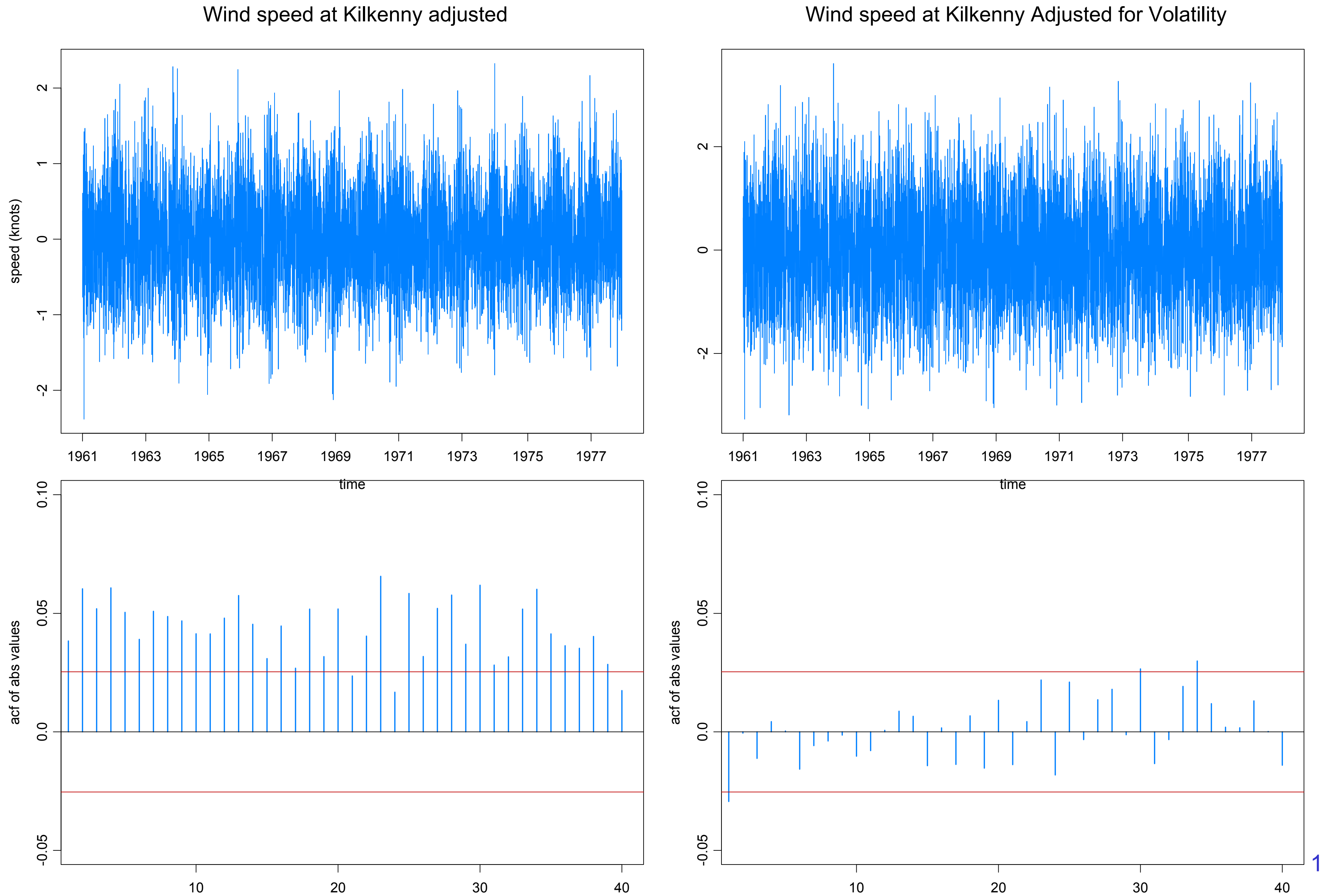


Time out: Illustration with ACF

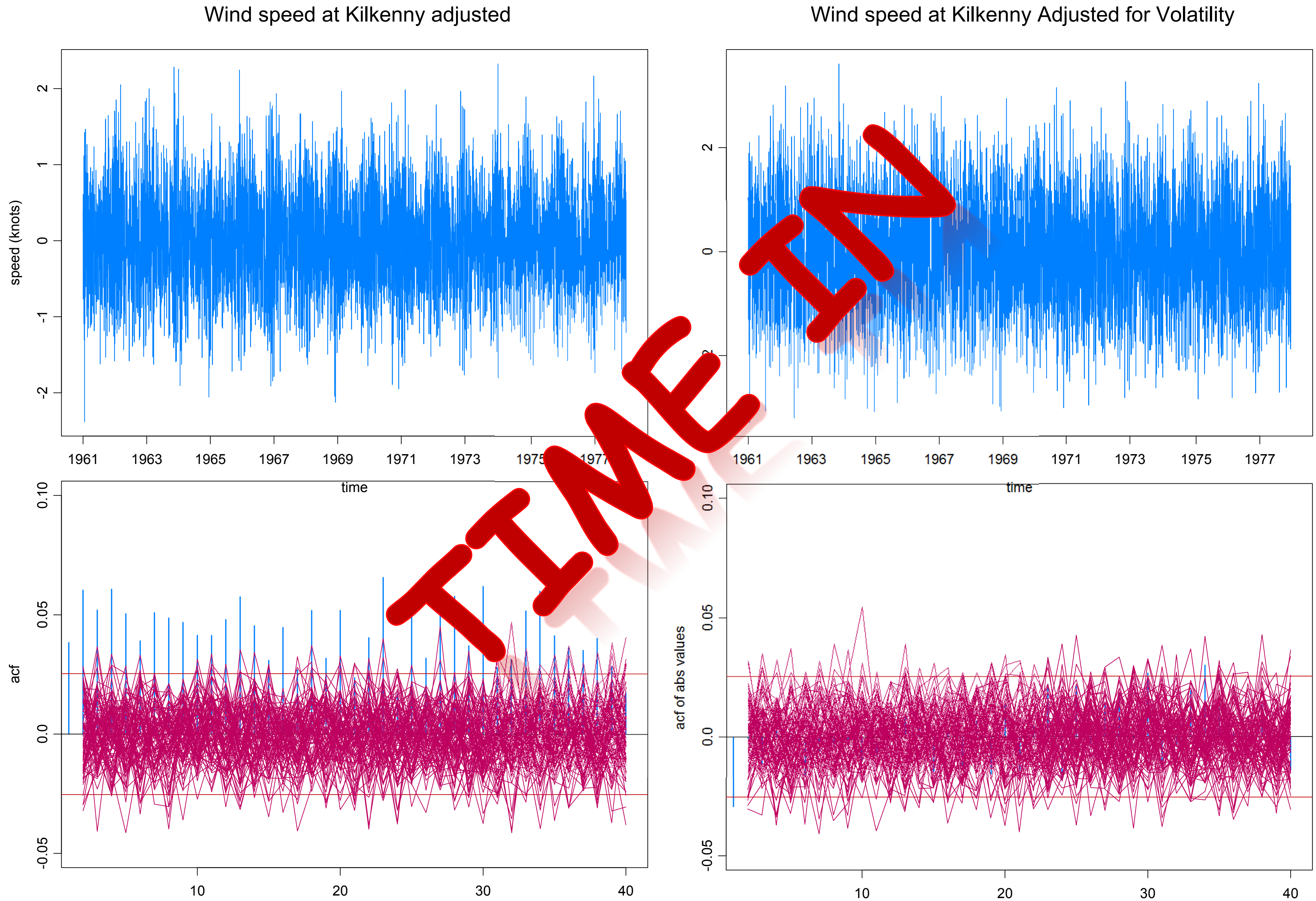
In plotting the sample ACF, one normally includes the ±1.96/sqrt(n) bounds (95% CI under the assumption of iid noise). One could use the permutation idea here as well.



Time out: Illustration with ACF



Time out: Illustration with ACF

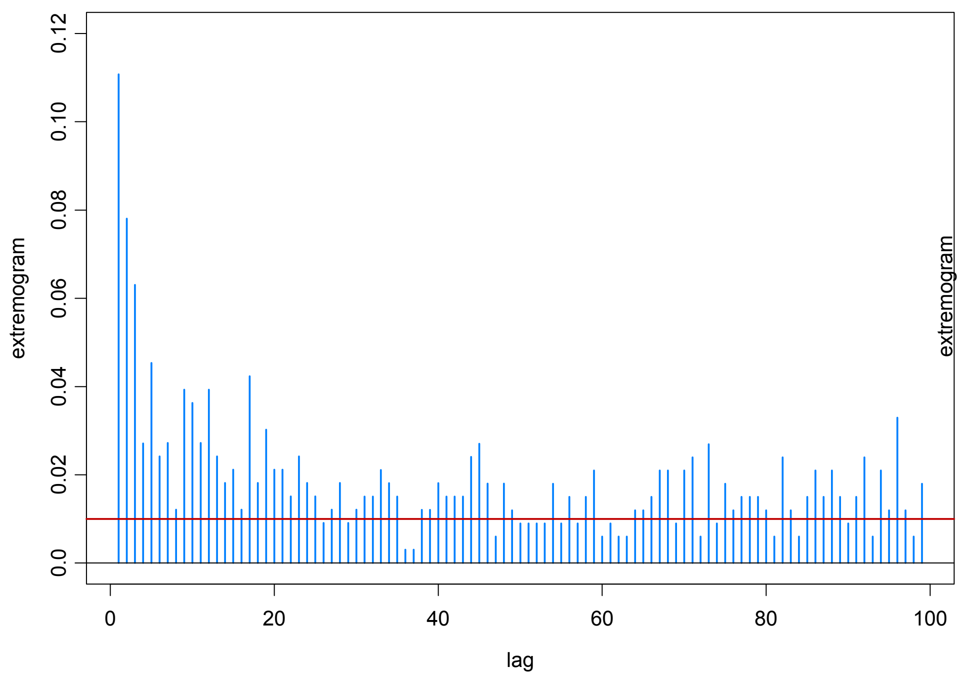


Application to Five-Minute Return Data (US/DM) exchange

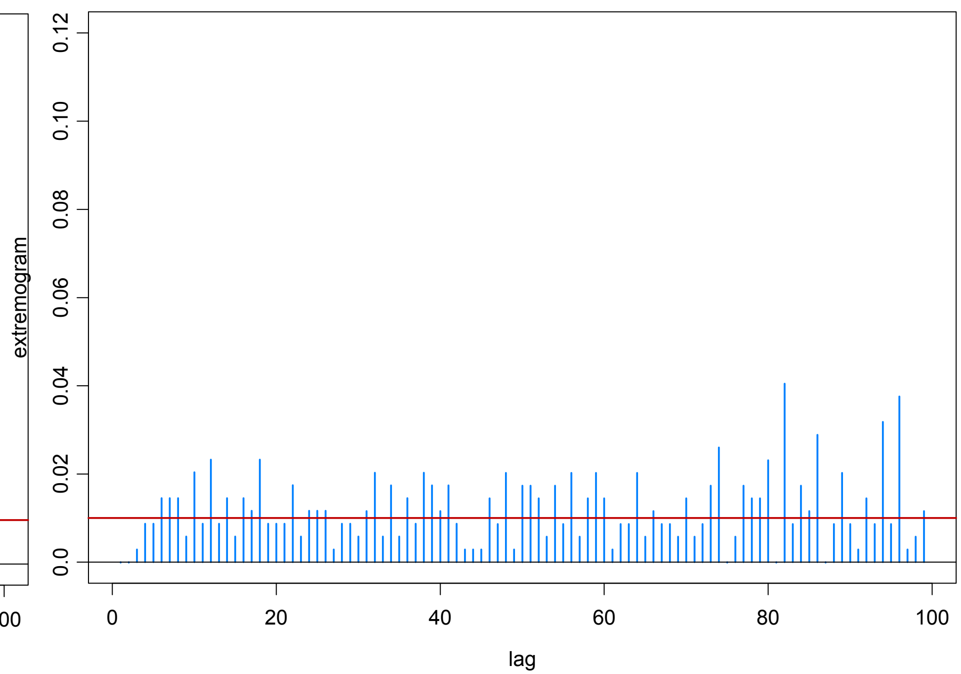
Extremogram for residuals from subset AR(18) and from GARCH

A=B=(1,∞)

Residuals from AR



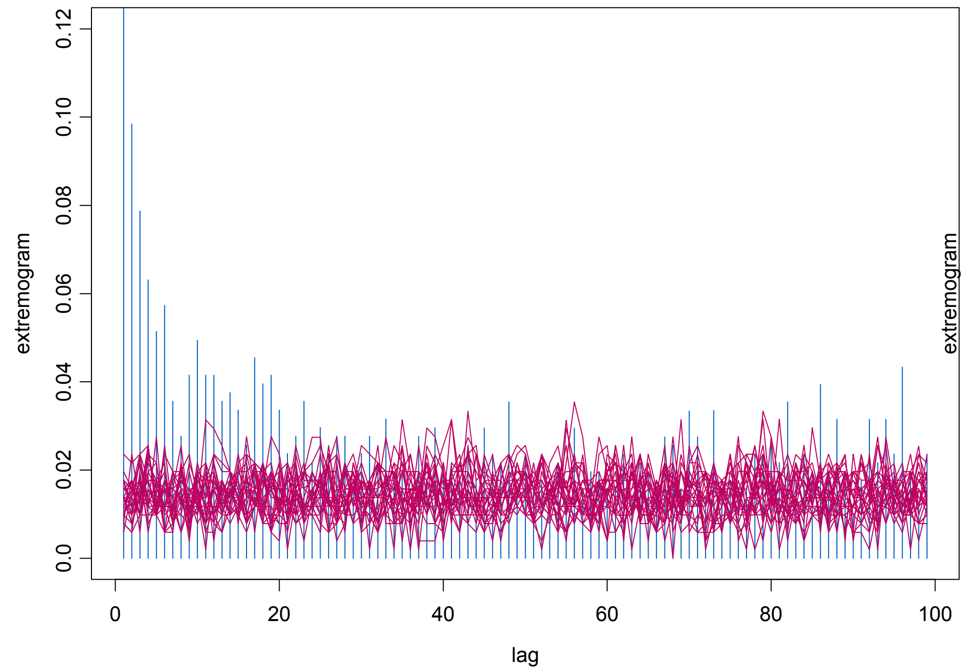
Residuals from GARCH



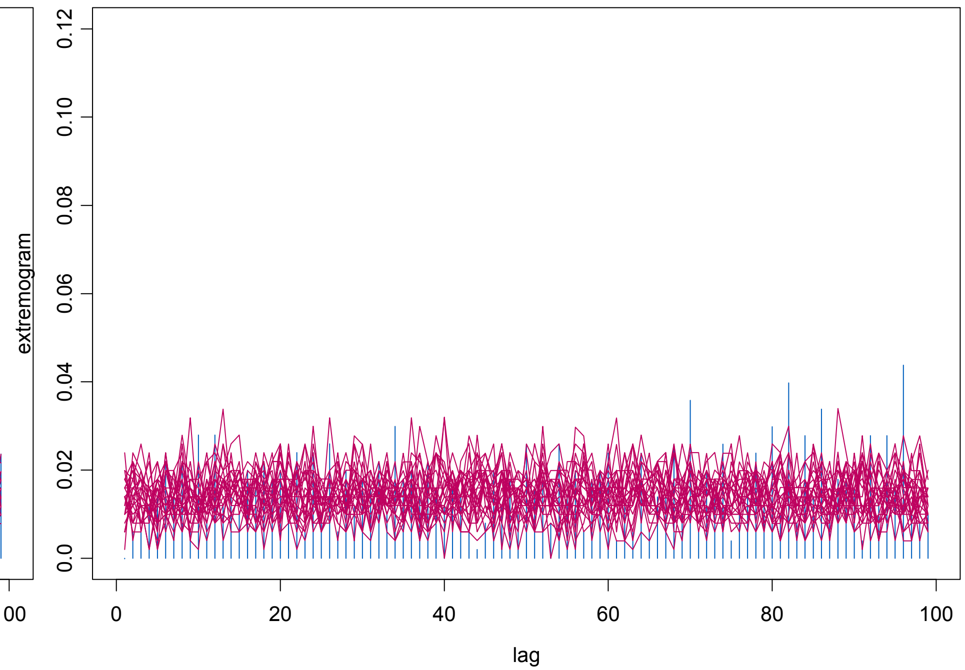
Application to Five-Minute Return Data (US/DM) exchange

Extremogram for residuals from subset AR(18) and from GARCH
A=B=(1,∞)

Residuals from AR



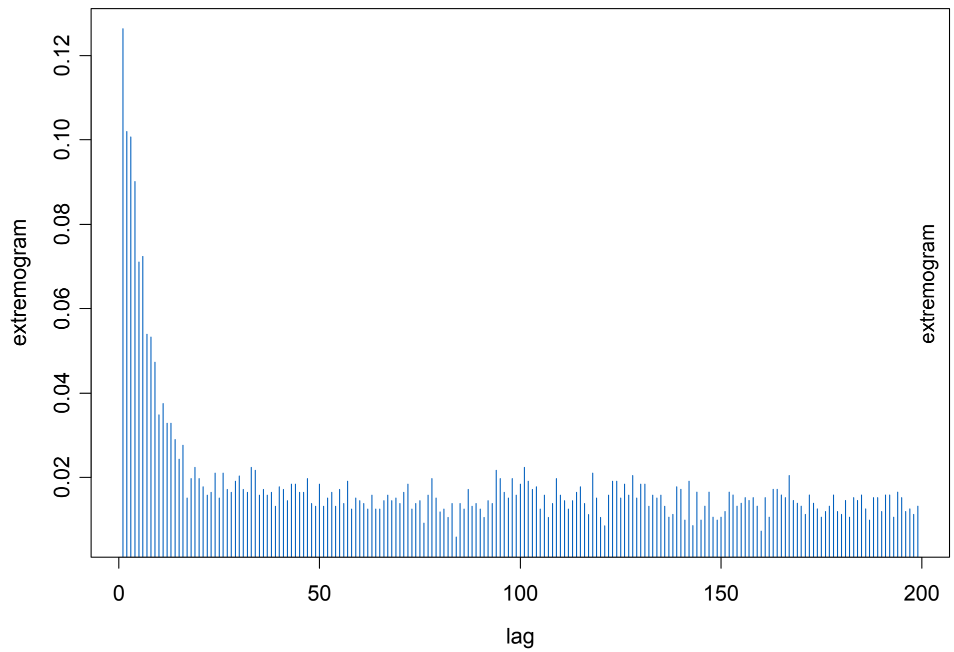
Residuals from GARCH

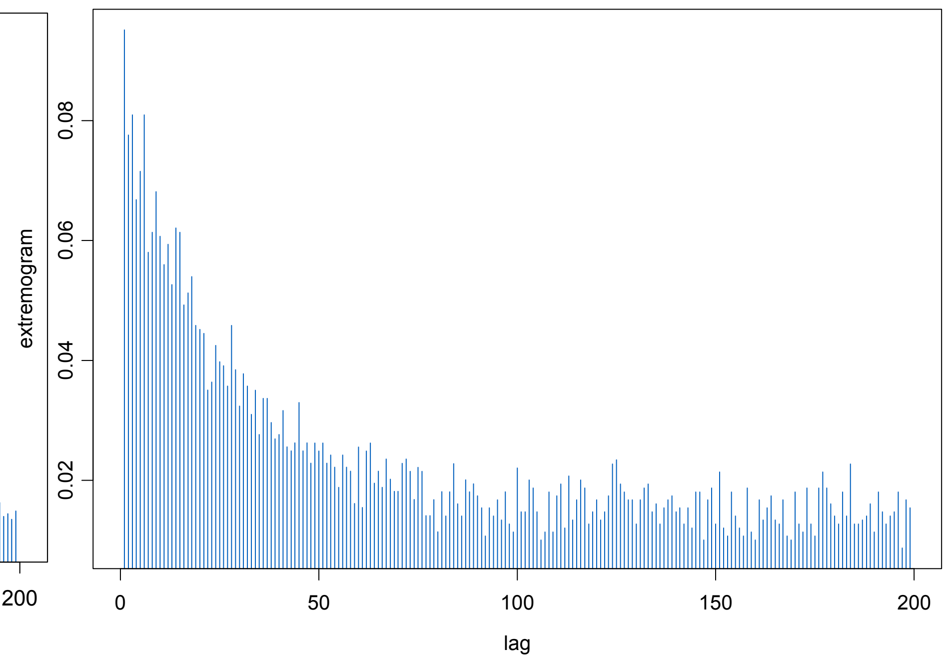


Extremogram of a SV Process

SV Process: $X_t=\sigma_t Z_t$, $\{Z_t\} \sim$ IID t_4 ; $\log \sigma_t = .9 \log \sigma_{t-1} + \varepsilon_t$

GARCH(1,1): $X_t=(.1+.14\, X^2_{t-1} +.83\sigma^2_{t-1})^{1/2} Z_t$, $\{Z_t\} \sim$ IID N(0,1),





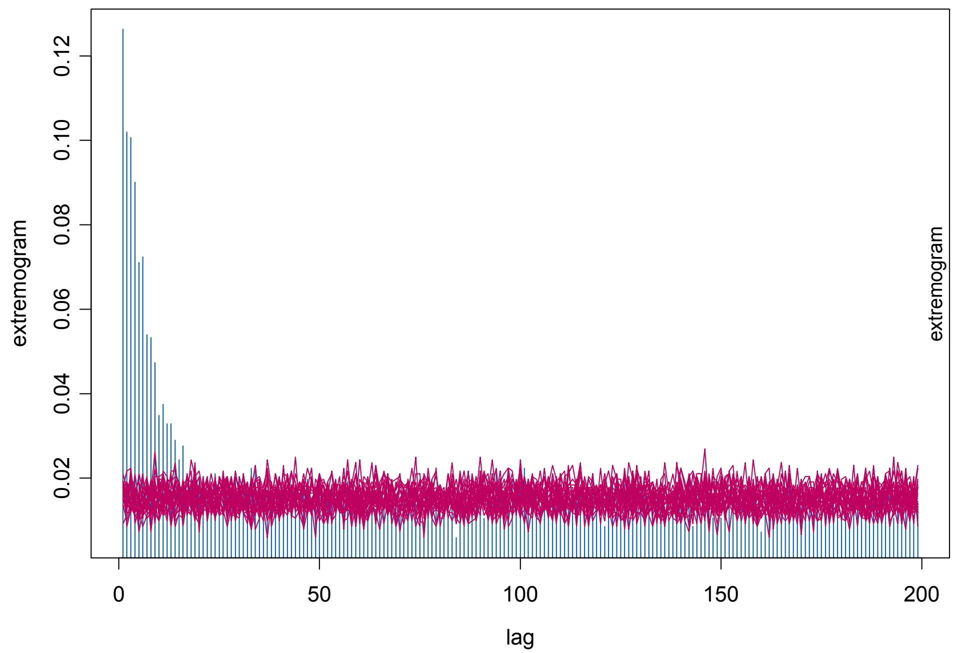
SV GARCH

Threshold = .97 quantile

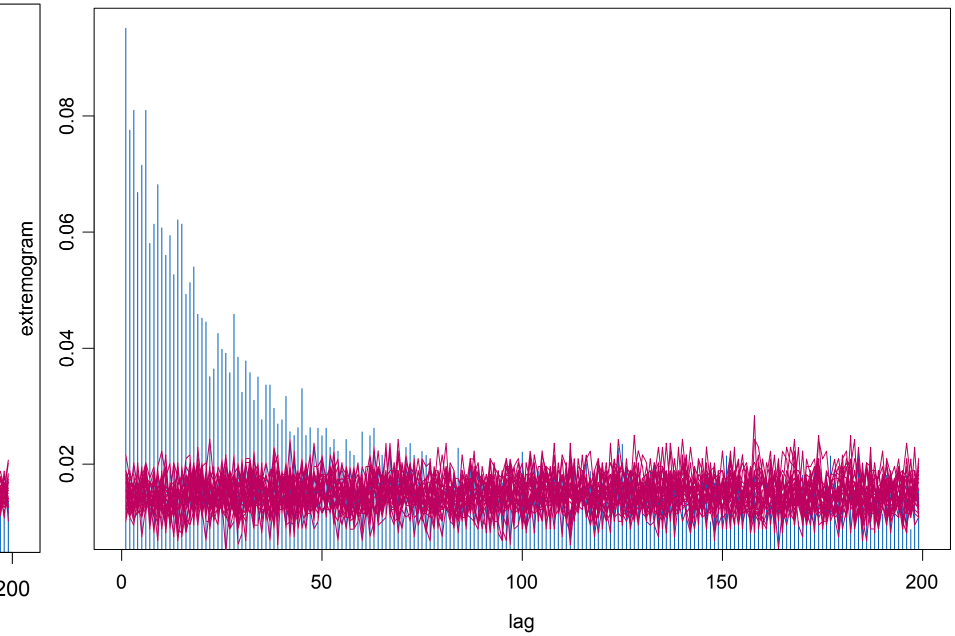
Extremogram of a SV Process

SV Process: $X_t = \sigma_t Z_t$, $\{Z_t\} \sim$ IID t_4 ; $\log \sigma_t = .9 \log \sigma_t + \epsilon_t$

GARCH(1,1): $X_t = (.1 + .14 X_{t-1}^2 + .83\sigma_t^2)^{1/2} Z_t$, $\{Z_t\} \sim$ IID N(0,1),



SV



GARCH

Threshold = .97 quantile

Extremogram of a Max-MA(2)

Example: Let (Z_t) be iid with Pareto distribution, i.e., $P(Z_1 > x) = x^{-\alpha}$ for $x \geq 1$, and set $X_t = \max(Z_t, Z_{t-1}, Z_{t-2})$. Then

$$nP(X_1 > xn^{1/\alpha}) \to 3x^{-\alpha} \quad \text{and} \quad F^n(xn^{1/\alpha}) \to \exp(-3x^{-\alpha}).$$

On the other hand,

$$P(n^{-1/\alpha} M_n \leq x) = P(n^{-1/\alpha} \max(Z_{-1}, \ldots, Z_n) \leq x) \to \exp(-x^{-\alpha}) = \exp(-1/3\ 3x^{-\alpha}),$$

which implies that the extremal index is $\theta = 1/3$.

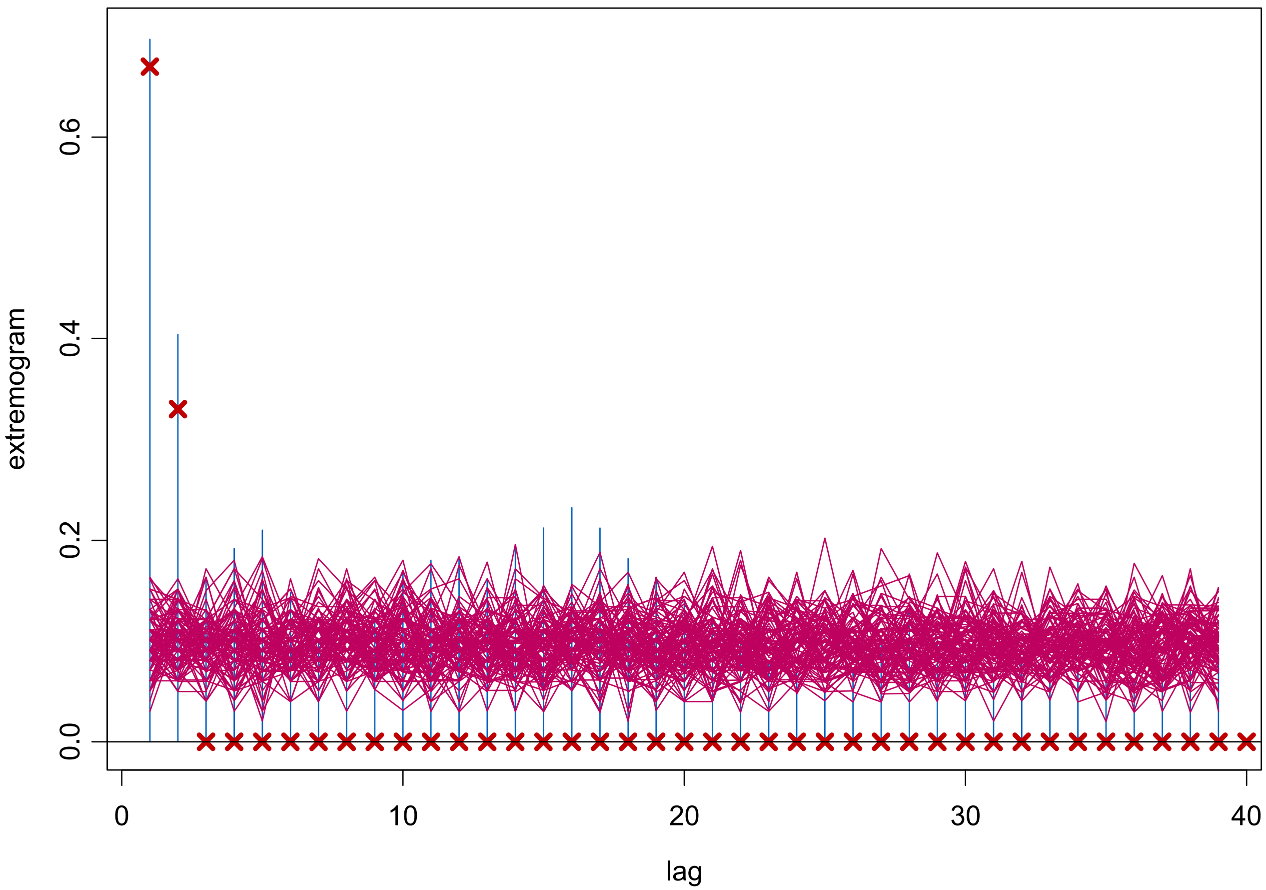
The extremogram with $A = B = (1, \infty)$ is

$$\begin{aligned} \lim_n P(X_h > n^{1/\alpha} \mid X_0 > n^{1/\alpha}) &= 1 \quad \text{for } h = 0. \\ &= 2/3 \quad \text{for } h = 1 \\ &= 1/3 \quad \text{for } h = 2 \\ &= 0, \quad \text{for } h > 2. \end{aligned}$$

Extremogram of a Max-MA(2)

Extremogram: $\lim_n P(X_h > n^{1/\alpha} \mid X_0 > n^{1/\alpha}) = 2/3, 1/3, 0$ for

$h = 1$, $h=2$, and for $h > 3$, respectively. Blue = sample

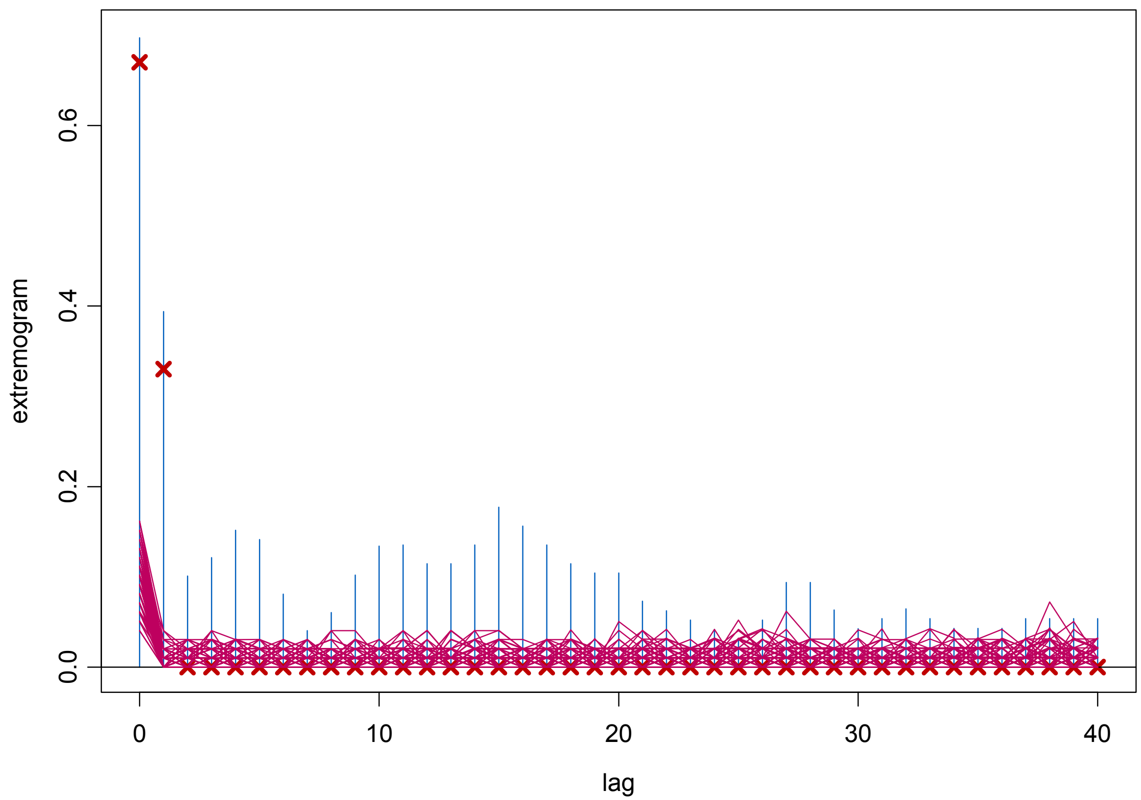


Extremogram of a Max-MA(2)

Extremogram: $\lim_n P(\min(X_h, X_{h+1}) > n^{1/\alpha} \mid X_0 > n^{1/\alpha}) = 2/3, 1/3, 0$ for

$h = 0$, $h=1$, and for $h > 2$, respectively.

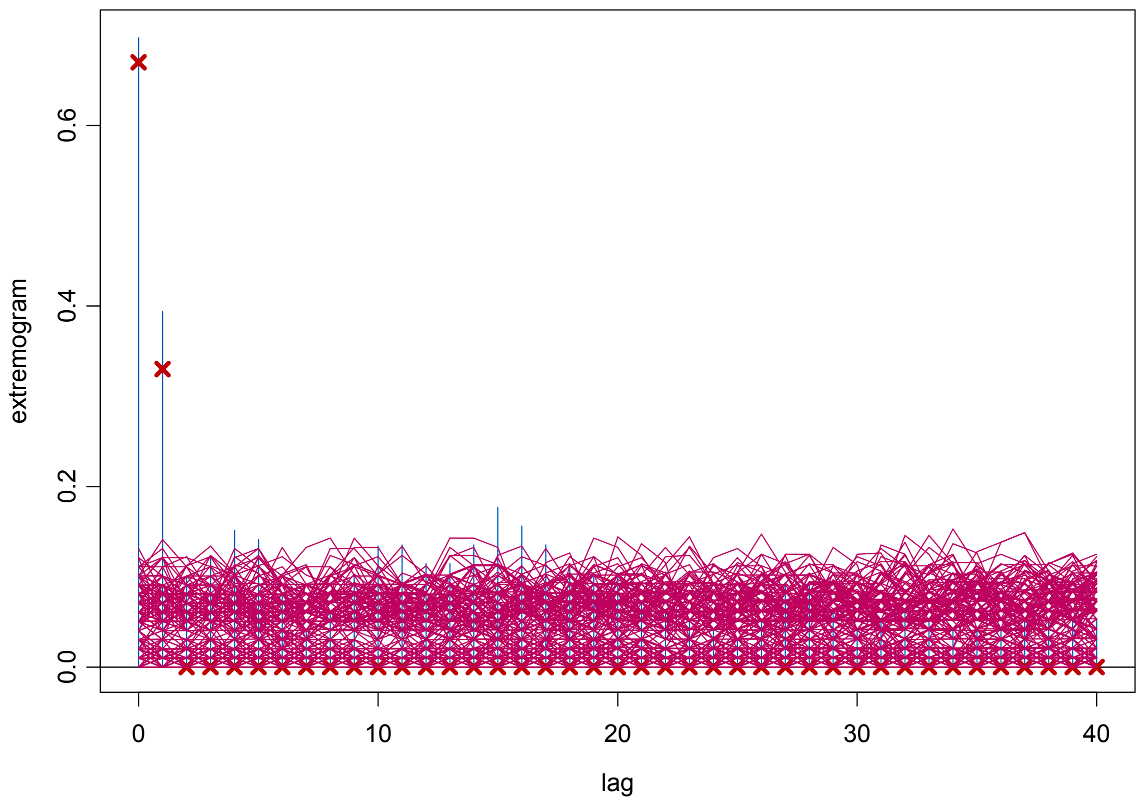
Note: Confidence intervals are narrow—how come?



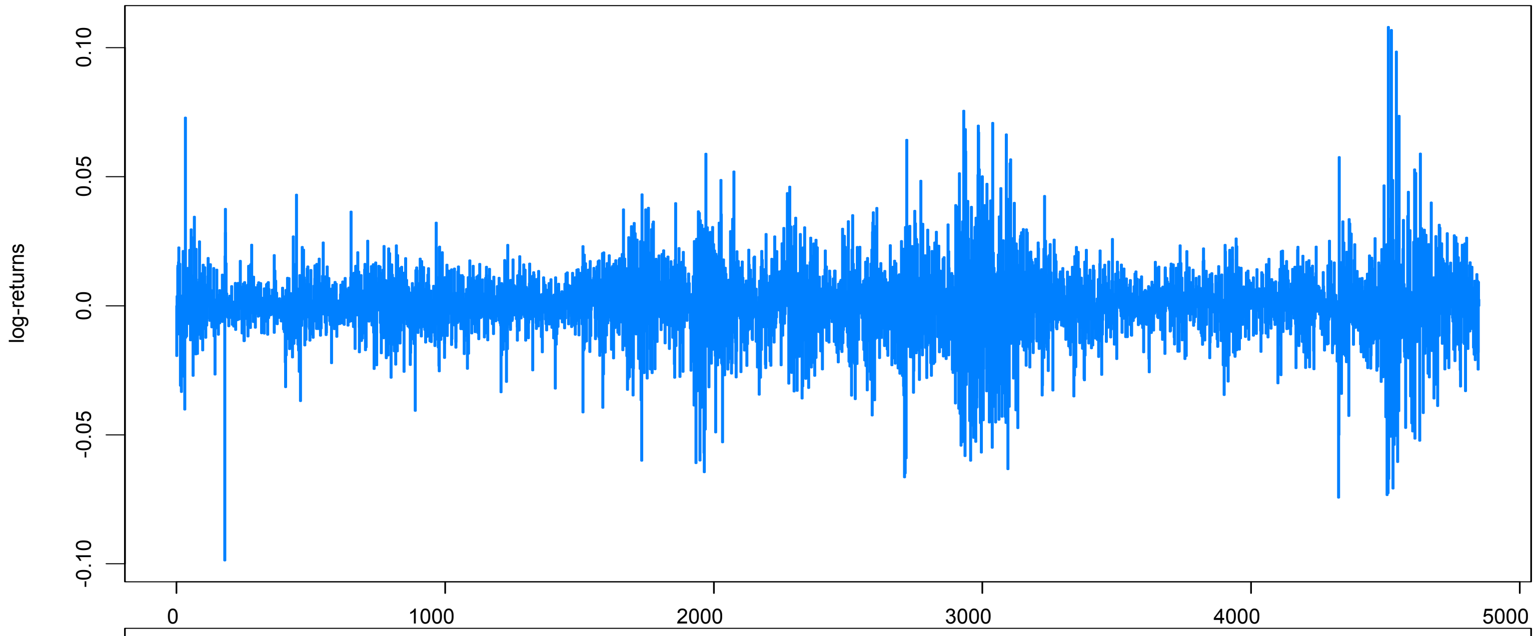
Extremogram of a Max-MA(2)

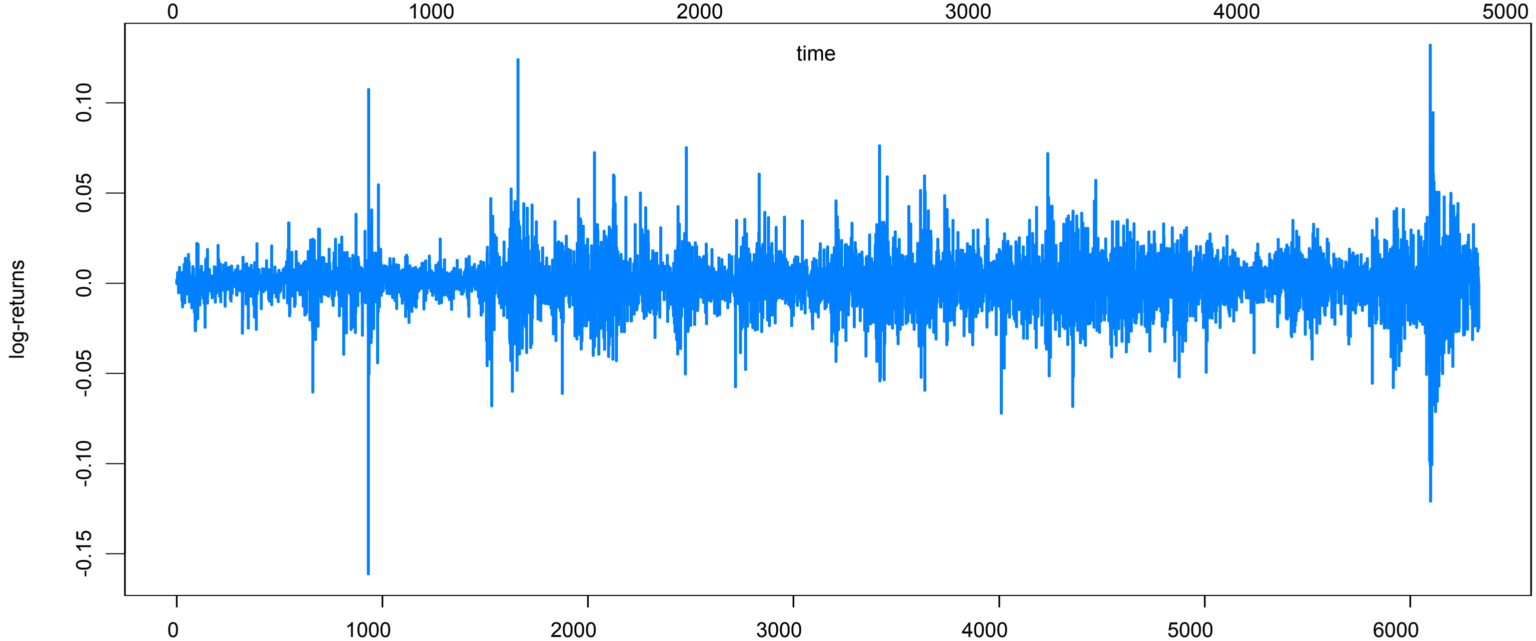
Extremogram: $\lim_n P(\min(X_h, X_{h+1}) > n^{1/\alpha} \mid X_1 > n^{1/\alpha}) = 2/3, 1/3, 0$ for $h = 0$, $h=1$, and for $h > 2$, respectively.

Note: Confidence intervals are narrow—how come?

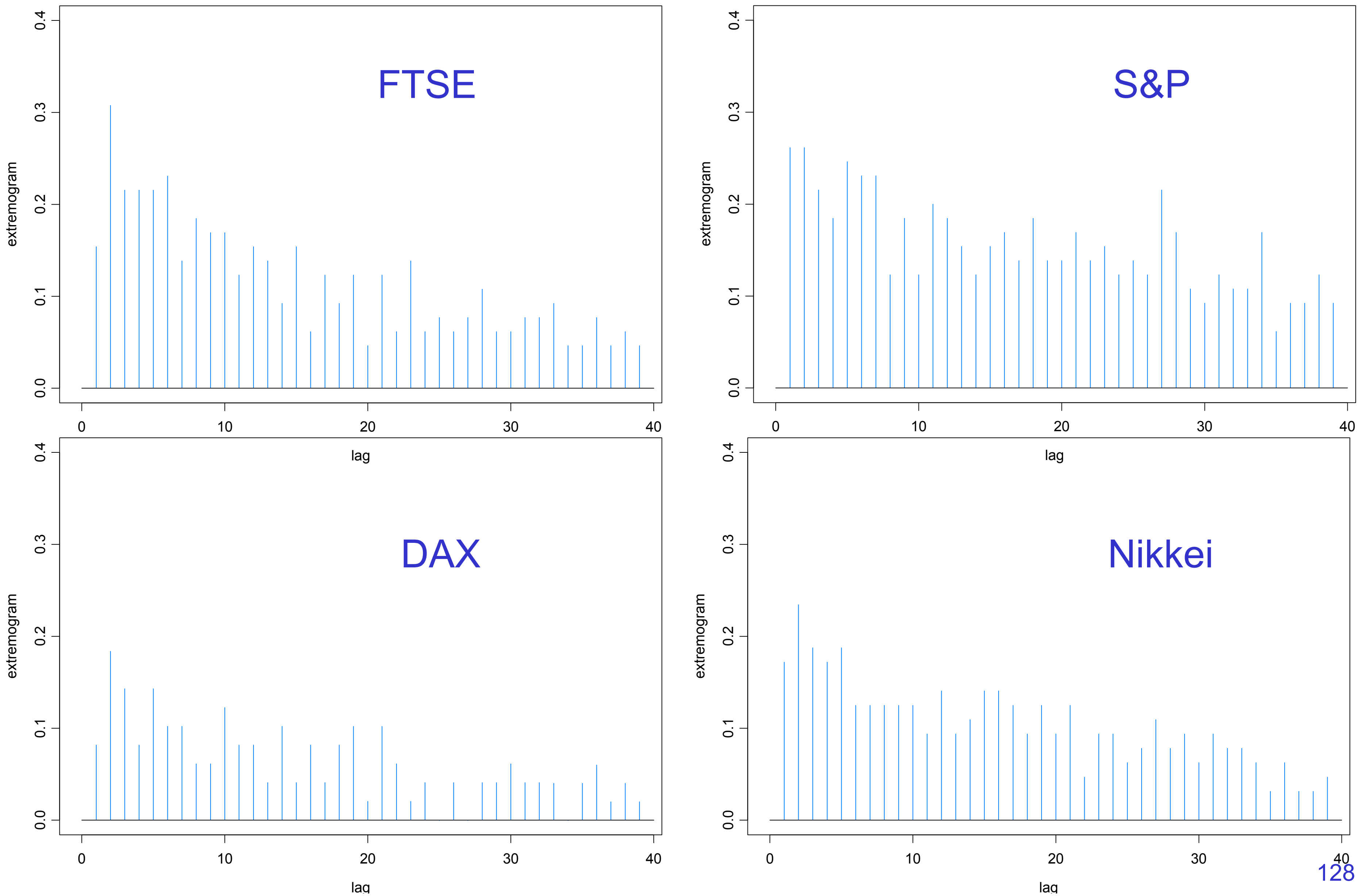


Log-returns for DAX and Nikkei (Apr 4, `84-Oct 2, `09

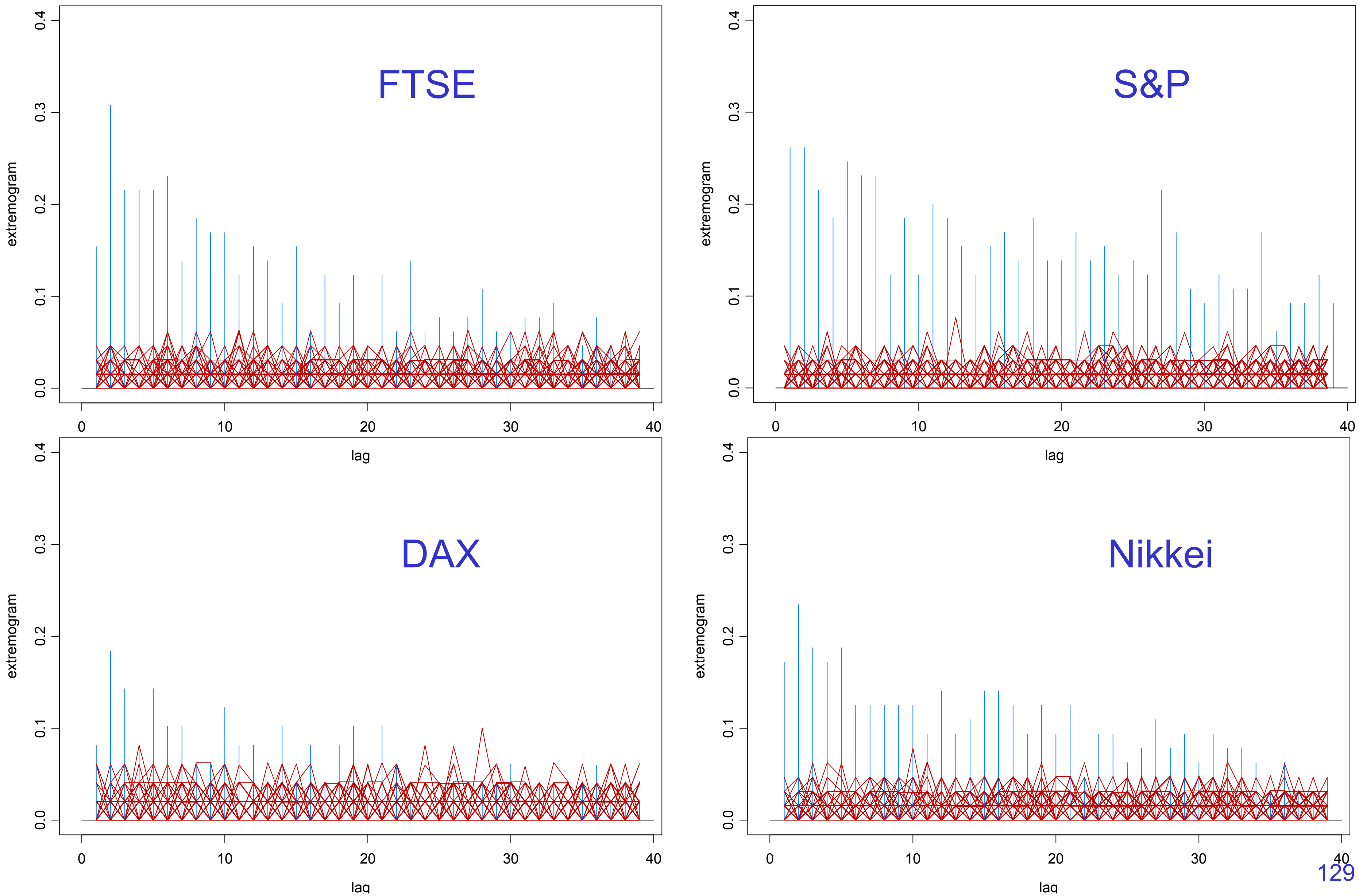




Extremogram for FTSE, S&P, DAX, Nikkei



Extremogram for FTSE, S&P, DAX, Nikkei



Cross-Extremogram

The cross-extremogram measures extremal dependence between two or more series. Suppose we have two time series $\{X_t\}$ and $\{Y_t\}$

Definition: For two sets A & B ***bounded away from 0***, the ***cross-extremogram*** is defined as

$$\rho_{A,B}(h) = \lim_{x\to\infty} P(Y_h \epsilon\, xB \mid X_0 \epsilon\, xA)$$

For example, if X_t and Y_t represent log-returns of two stocks, then one might be interested in extremal dependence of negative returns. It may seem natural to take $A = B = (-\infty, -1]$, so that

$$\rho_{A,B}(h) = \lim_{x\to\infty} P(Y_h < -x \mid X_0 < -x).$$

Cross-Extremogram

As before, we estimate

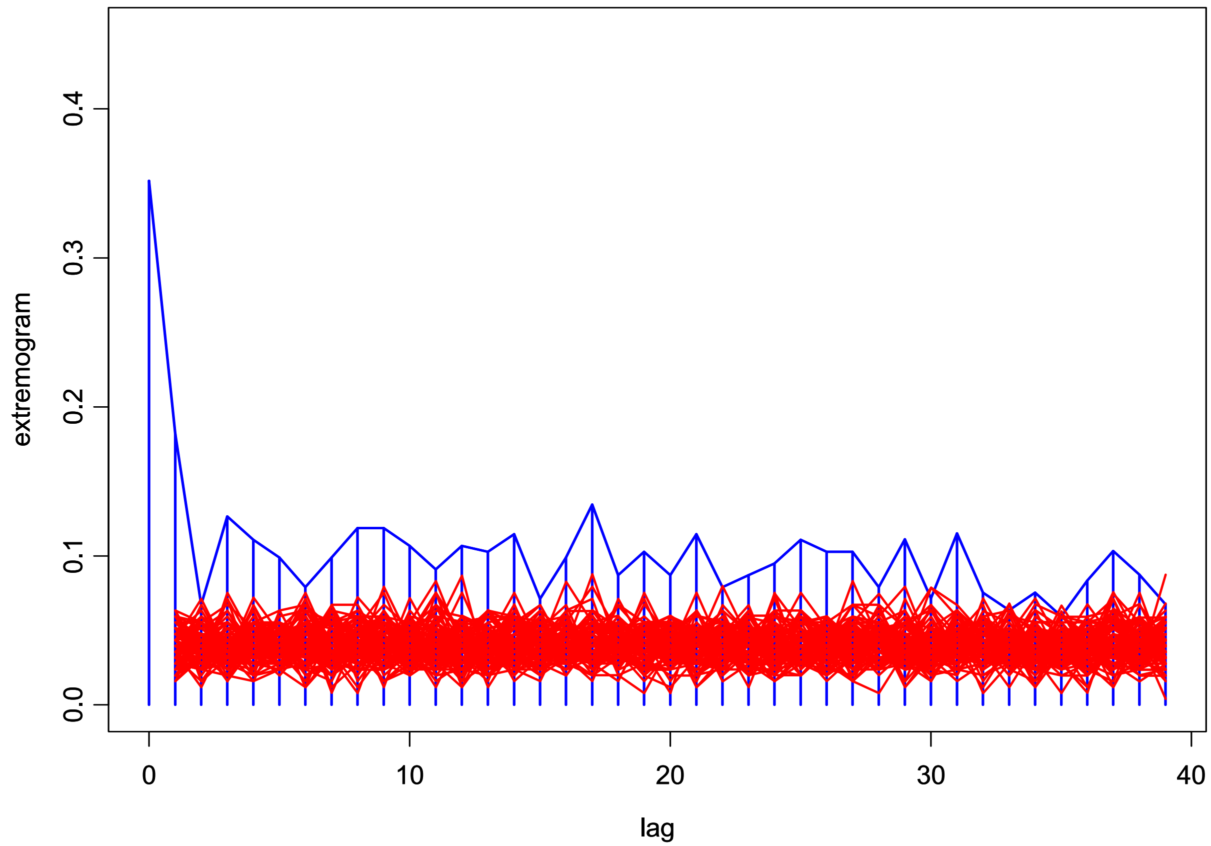
$$\rho_{A,B}(h) = \lim_{x\to\infty} P(Y_h \epsilon\, xB \mid X_0 \epsilon\, xA)$$

by

$$\hat{\rho}_{A,B}(h) = \frac{\frac{m}{n}\sum_{t=1}^{n-h} I_{\{a_{m,1}^{-1}X_t \in A,\, a_{m,2}^{-1}Y_{t+h} \in B\}}}{\frac{m}{n}\sum_{t=1}^{n} I_{\{a_{m,1}^{-1}X_t \in A\}}}$$

Problem: For log-returns, heteroskedasticity can produce ***spurious*** extremograms. That is, volatility in both series (which tend to happen in unison) produce large extremograms.

Cross-Extremogram FTSE and SP

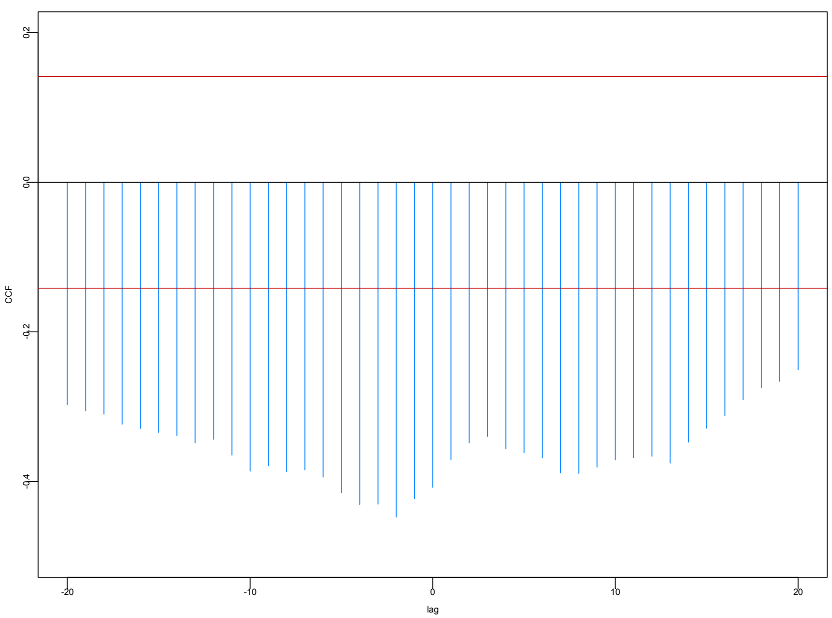


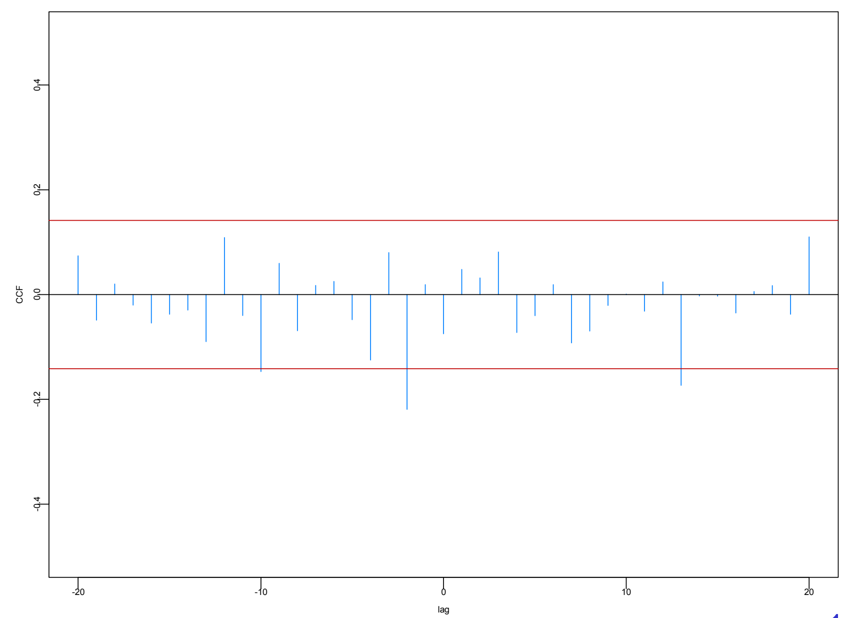
Cross-Extremogram

Strategy: Devolatilize the component series before computing the extremogram. This is ***analogous*** to the issue of spurious cross-correlations in a time series without whitening the series first.

Cross-correlation between two "independent" AR(1)'s

Cross-correlation between the ***whitened*** series'





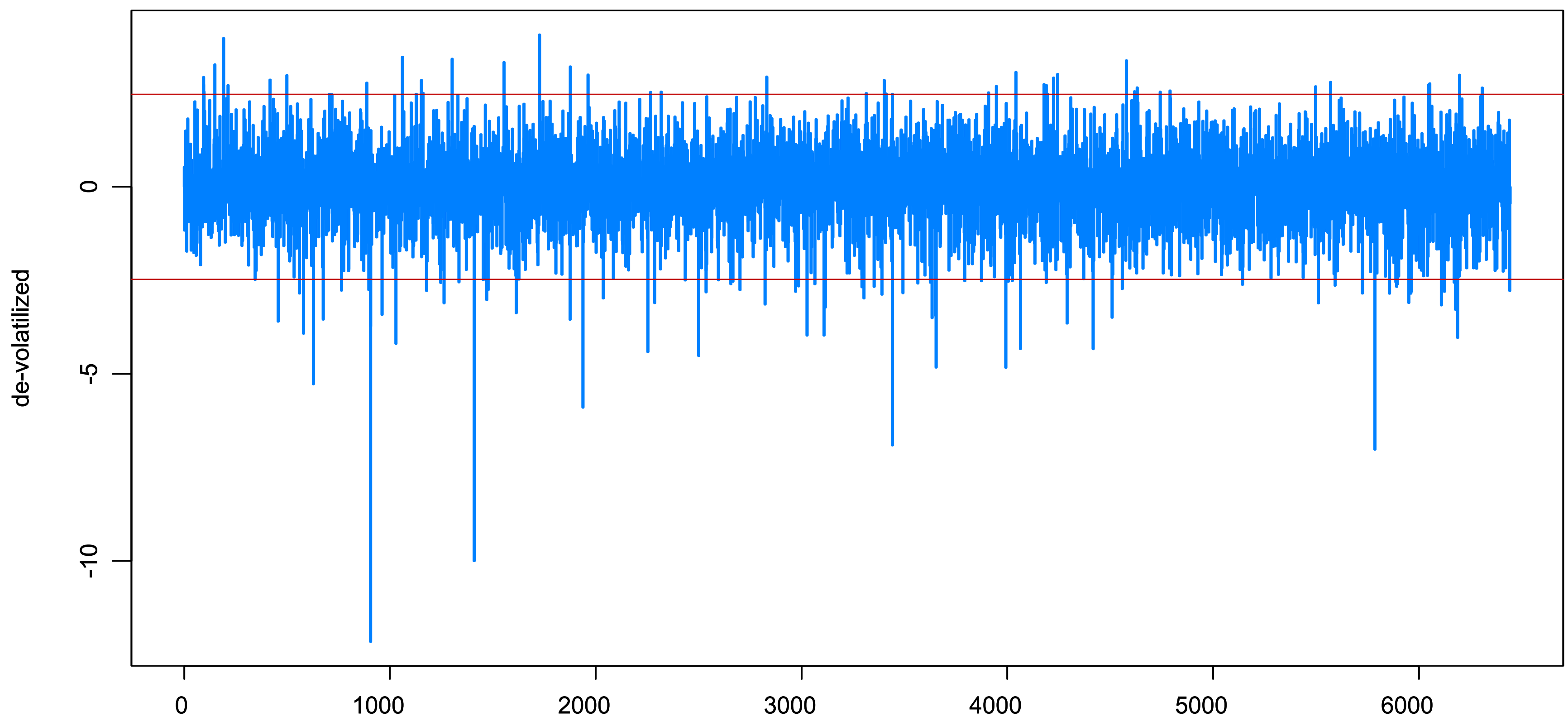
Devolatilizing (deGARCHing) S&P

Extremogram for S&P: significant for large number of lags ~40+

Devolatilize S&P by fitting GARCH(1,1):

$$X_t = (6.28e-7 + .057X_{t-1}^2 + .939\sigma_{t-1}^2)^{1/2} Z_t,$$

$\{Z_t\} \sim IID\ t(6.14),$



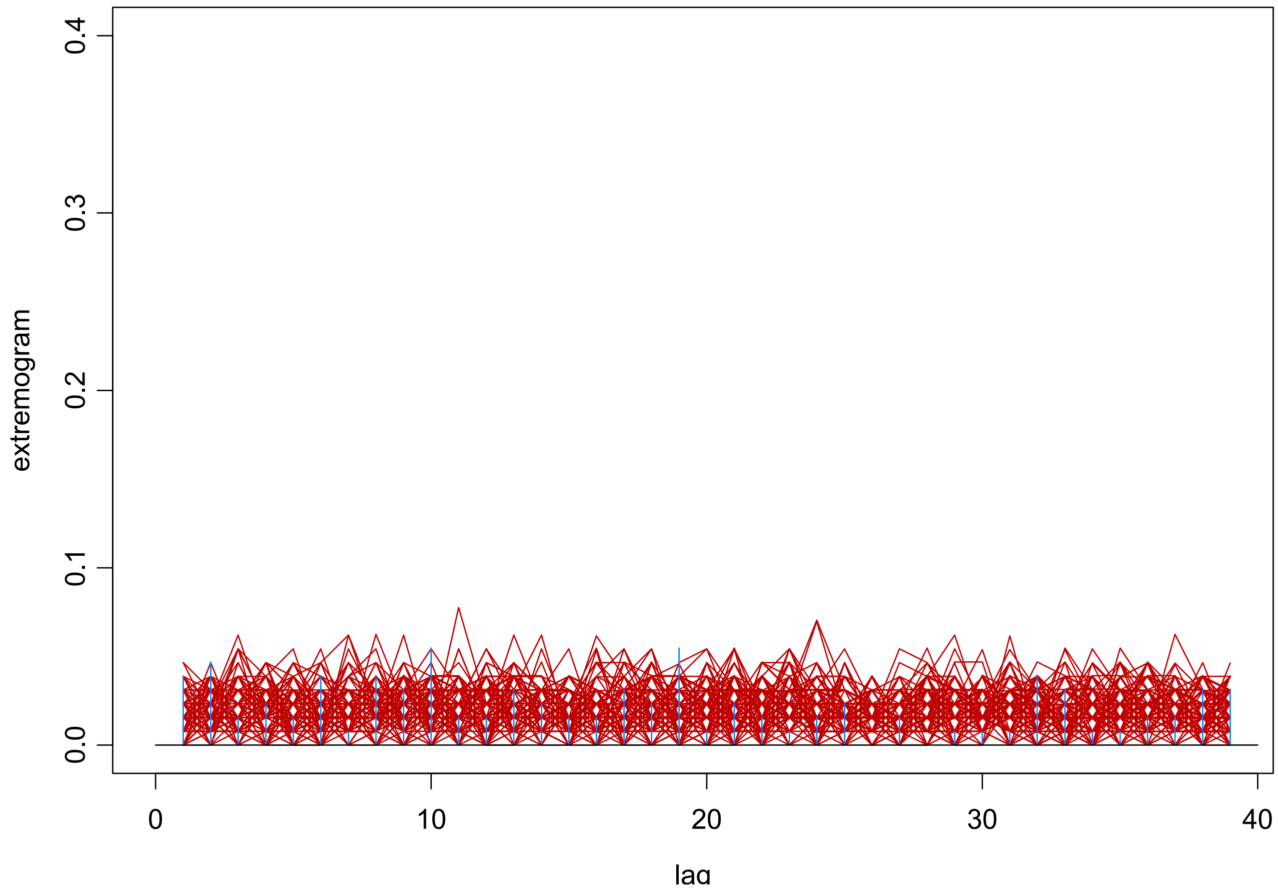
Devolatilizing S&P

Extremogram for S&P: significant for large number of lags ~40+

Devolatilize S&P by fitting GARCH(1,1):

$$X_t = (6.28e-7 + .057X_{t-1}^2 + .939\sigma_{t-1}^2)^{1/2} Z_t,$$

$\{Z_t\} \sim IID\ t(6.14),$

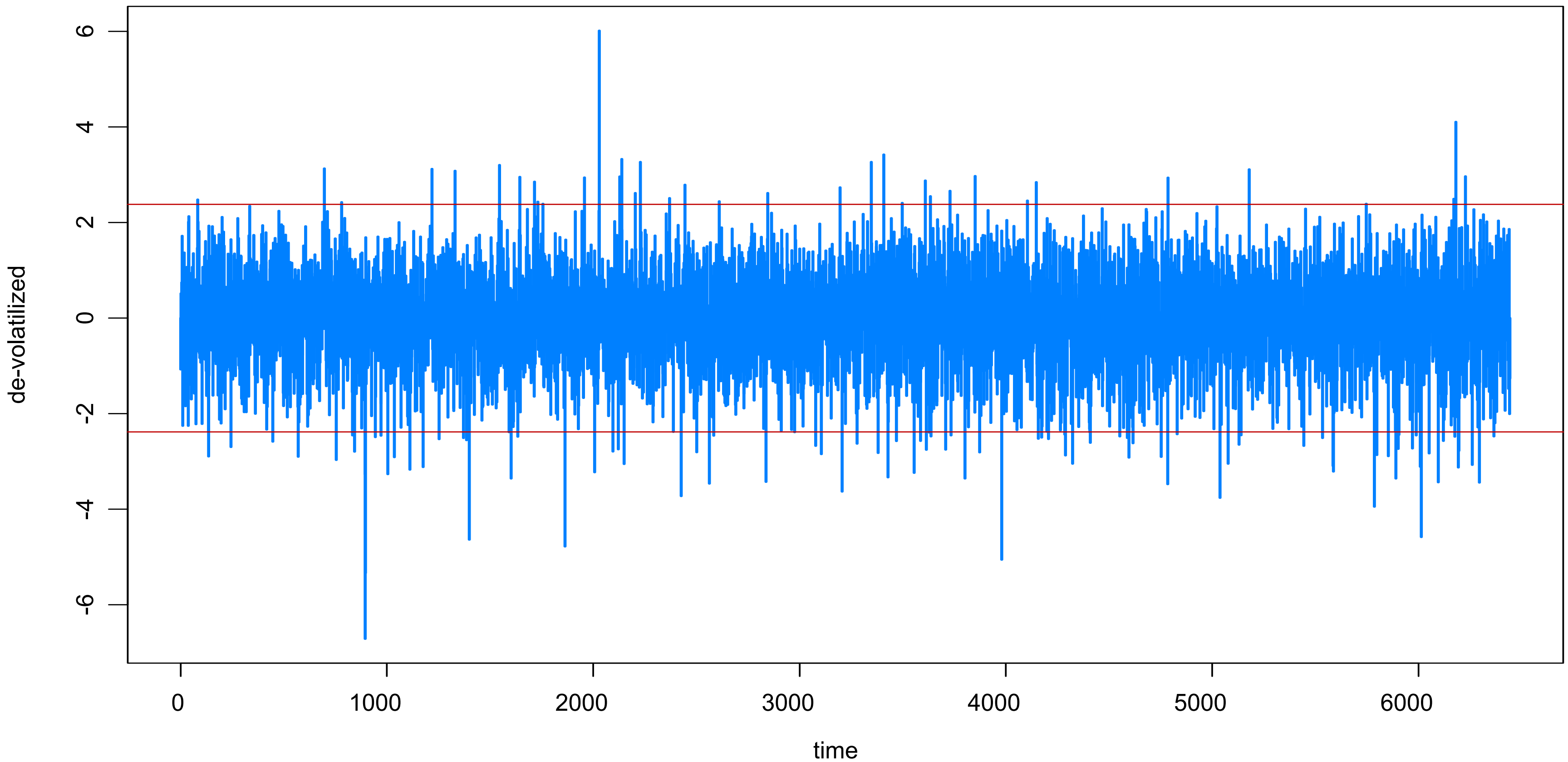


Devolatilizing (deGARCHing) FTSE

Extremogram for FTSE: significant for large number of lags ~40+

Devolatilize FTSE by fitting GARCH(1,1):

$$X_t=(1.32e{-}6+.084\, X^2_{t-1} + .904\sigma^2_{t-1})^{1/2} Z_t, \quad \{Z_t\} \sim \text{IID } t(13),$$

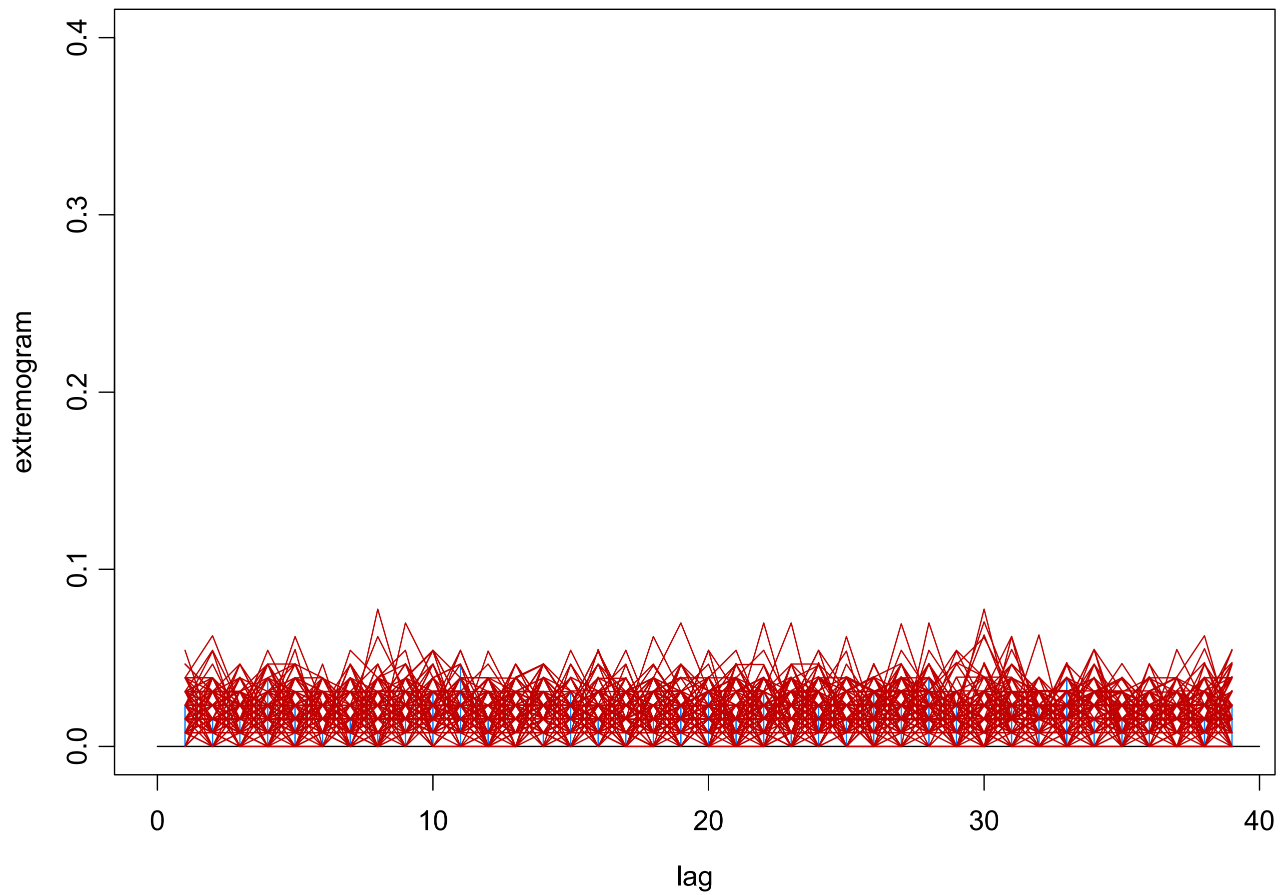


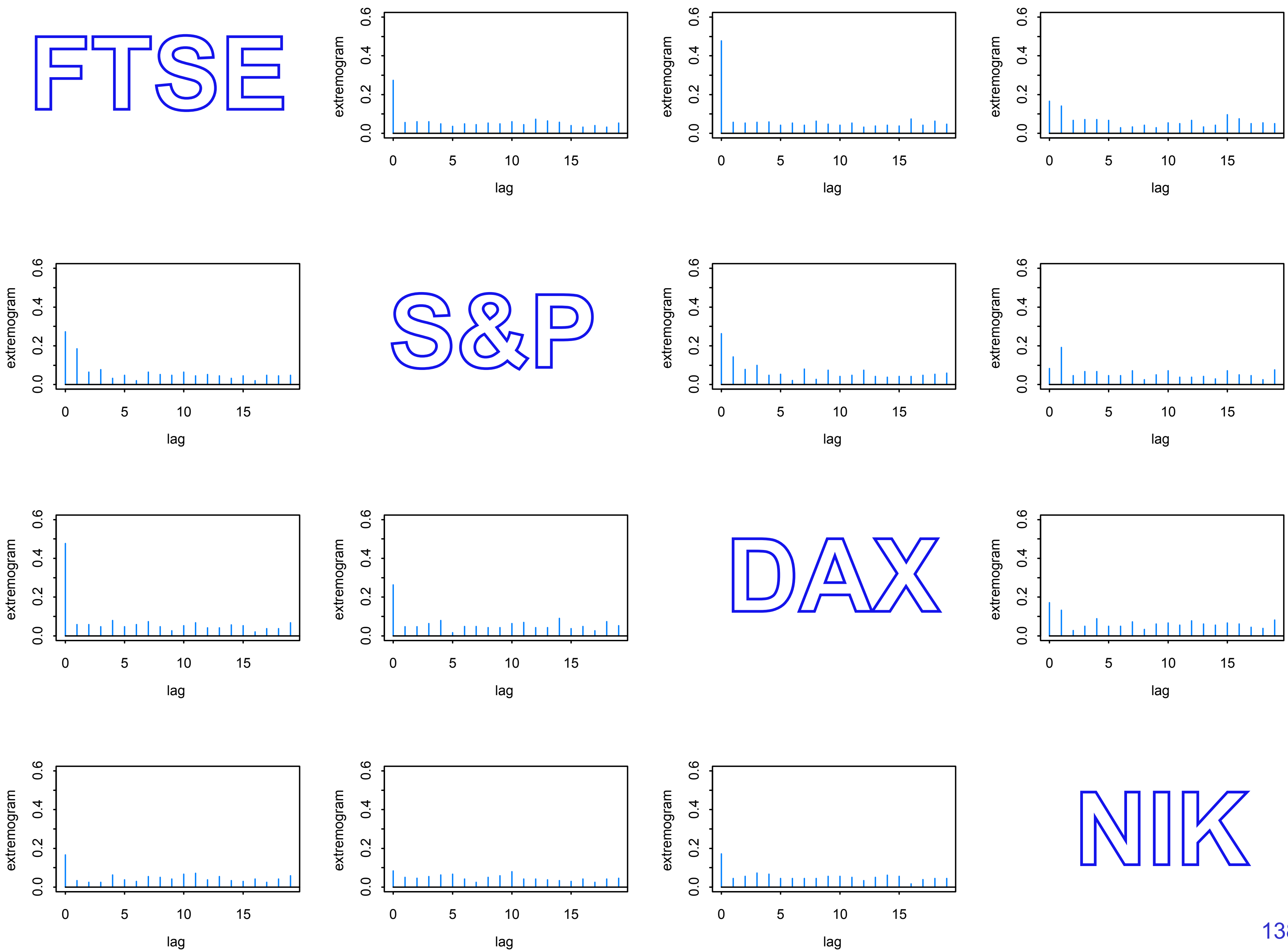
Devolatilizing FTSE

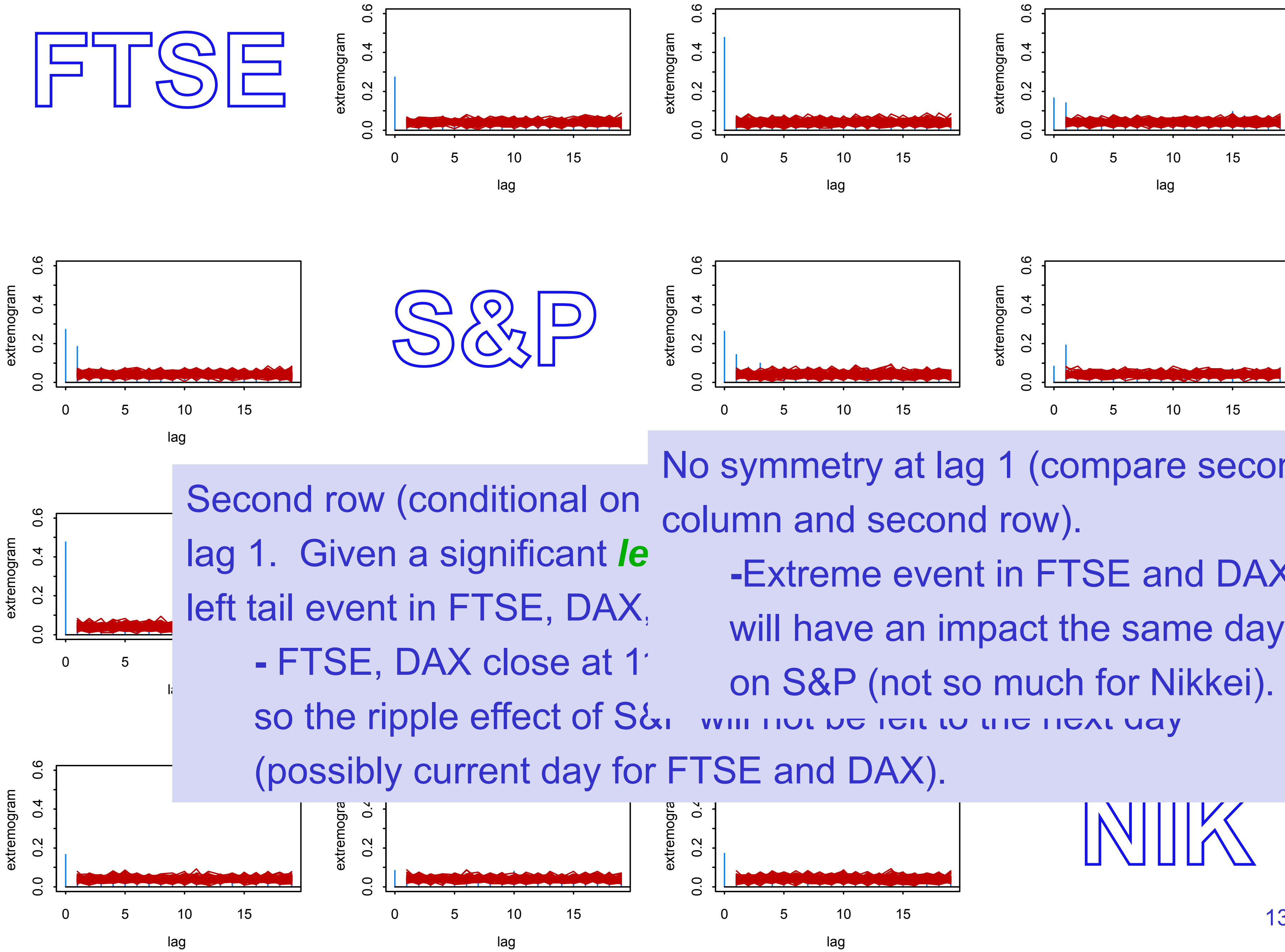
Extremogram for FTSE: significant for large number of lags ~40+

Devolatilize FTSE by fitting GARCH(1,1):

$$X_t=(6.28\text{e-}7+.057\,X^2_{t-1}+.939\sigma^2_{t-1})^{1/2}Z_t, \quad \{Z_t\} \sim \text{IID } t(6.14),$$




FTSE
S&P
DAX
NIK
extremogram
lag
0.0
0.2
0.4
0.6
0
5
10
15



Second row (conditional on
lag 1. Given a significant ***le***
left tail event in FTSE, DAX,
- FTSE, DAX close at 1
so the ripple effect of S&
(possibly current day for FTSE and DAX).

No symmetry at lag 1 (compare second column and second row).

-Extreme event in FTSE and DAX will have an impact the same day on S&P (not so much for Nikkei).

Cross-Extremogram for 3 Time Series

We extend the cross-extremogram to 3 time series.

Definition: For three sets A, B & C ***bounded away from 0***, the ***cross-extremogram*** is defined as

$$\rho_{A,B,C}(h) = \lim_{x\to\infty} P(Z_h \in xC, Y_h \in xB \mid X_0 \in xA)$$

We estimate $\rho_{A,B,C}(h)$ as before the empirical cross-extremogram.

To illustrate, we will look at 5 min log-returns , Dec 1, \`04-July 26, \`06.

X_t = 5 minute log-returns Bank of America

Y_t = 5 minute log-returns Citibank

Z_t = 5 minute log-returns Microsoft

Two cases:

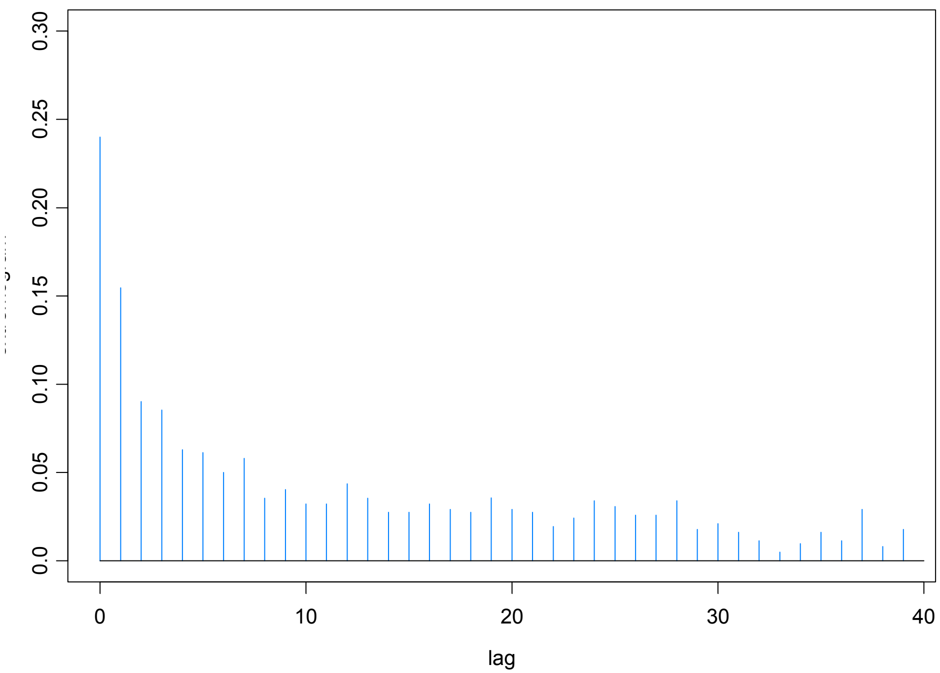
(i) $\rho_{A,B,C}(h) = \lim_{x\to\infty} P(|Z_h| > x \text{ or } |Y_h| > x \mid |X_0| > x)$

(ii) $\rho_{A,B,C}(h) = \lim_{x\to\infty} P(|Z_h| > x \mid |Y_0| > x \text{ or } |X_0| > x)$

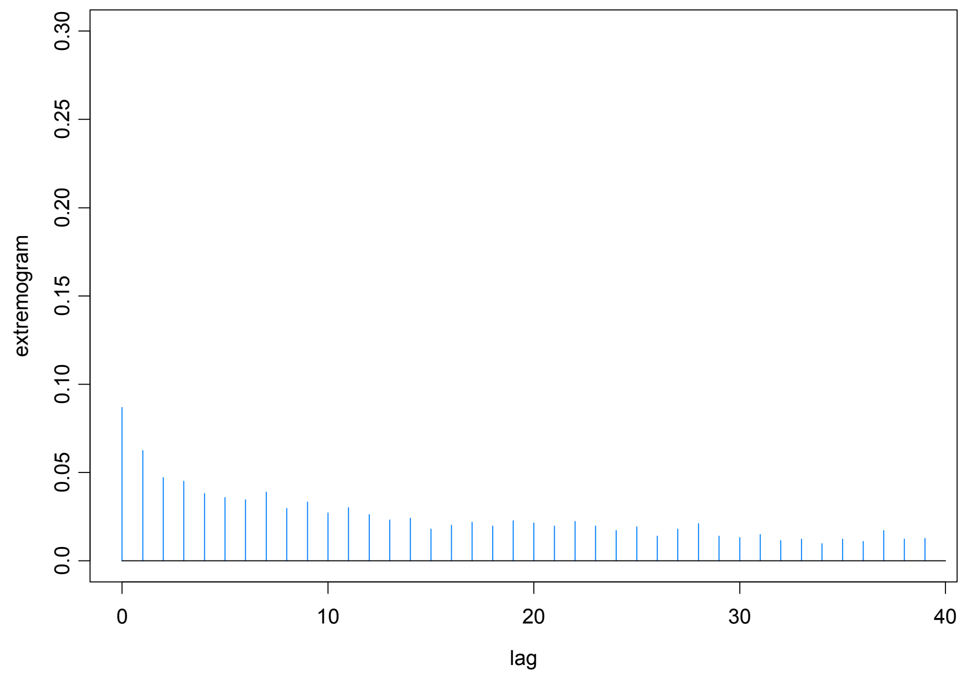
Cross-Extremogram for 3 Time Series

X_t = BAC, Y_t = Citibank, Z_t = MSFT

(i) $P(|Z_h| > x \; or \; |Y_h| > x \mid |X_0| > x)$



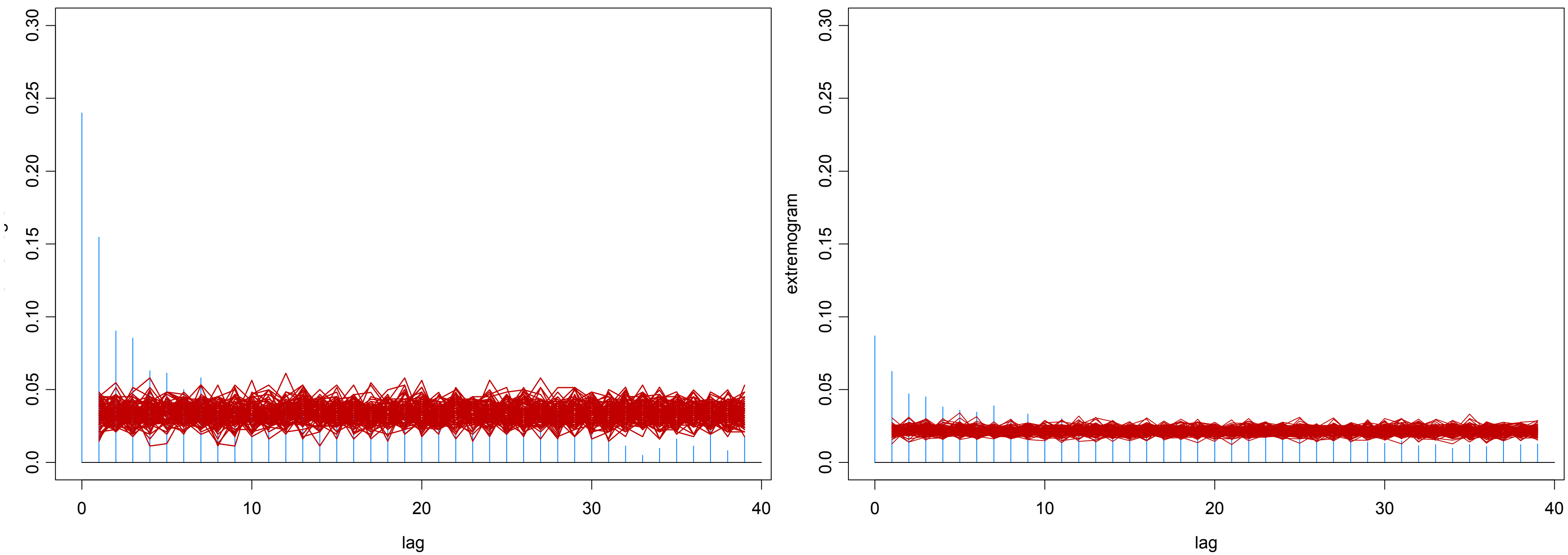
(ii) $P(|Z_h| > x \mid |Y_0| > or \; |X_0| \mid > x)$



Cross-Extremogram for 3 Time Series

X_t = BAC, Y_t = Citibank, Z_t = MSFT

(i) $P(|Z_h| > x \; or \; |Y_h| > x \mid |X_0| > x)$ (ii) $P(|Z_h| > x \mid |Y_0| > or \; |X_0| \mid > x)$



(i) Given BAC is large ⇒ CB or MSFT is large in same time period or 5 minutes hence.

(ii) Given BAC or CB large ⇒ MSFT is large in one time period.

Connections with Return Times (of rare events)

This is an idea due to ***Geman and Chang (2009)*:**

Setup:

- $\{X_t\}$ time series—think log-returns, for example.
- ξ_v, ξ_{1-v} are the vth and (1-v)th quantile of the of the marginal distribution.

Define the exceedance (or stopping times) times τ_j by

$$\tau_1 = \min\{ t \geq 1: X_t < \xi_v \text{ or } X_t < \xi_{1-v}\}$$

$$\tau_{j+1} = \min\{ t \geq \tau_j: X_t < \xi_v \text{ or } X_t < \xi_{1-v}\}\ ,\ j \geq 0.$$

The inter-arrival (or return times) are

$$T_j = \tau_j - \tau_{j-1} \quad j \geq 1.$$

These are the times between occurrences of rare events (number of tosses of a coin until next head).

Connections with Return Times (of rare events)

For ***nice*** time series, like iid observations, the T_j's are iid with a geometric distribution,

$$P(T_j = k) = (1-p)^{k-1}p, \quad k=1,2, \ldots,$$

$$p = P(X_t < \xi_v \text{ or } X_t > \xi_{1-v}) = 2v.$$

Recall for a geometric rv,

$$E(T_1) = 1/p.$$

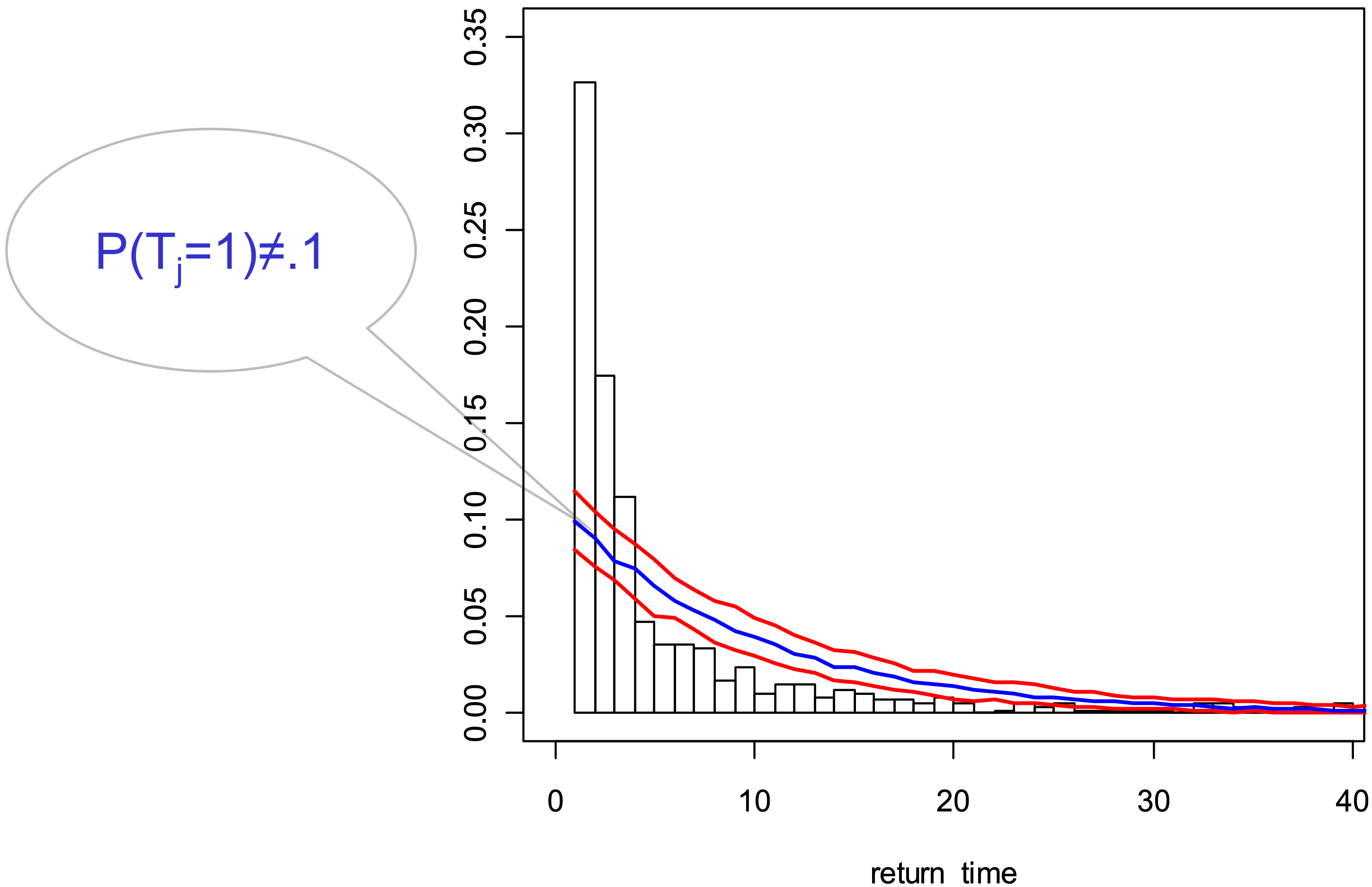
Note: This is the ***backstory*** behind the term 100 year flood, or 100 year *blank*, which corresponds to the threshold x such that the expected time until x is exceeded is 100. (In this case, $p = .01$, $x = \xi_{.99}$.)

Idea: For v fixed (can do one-sided tail), look at the histogram of return times and compare against a geometric distribution.

Connections with Return Times (of rare events)

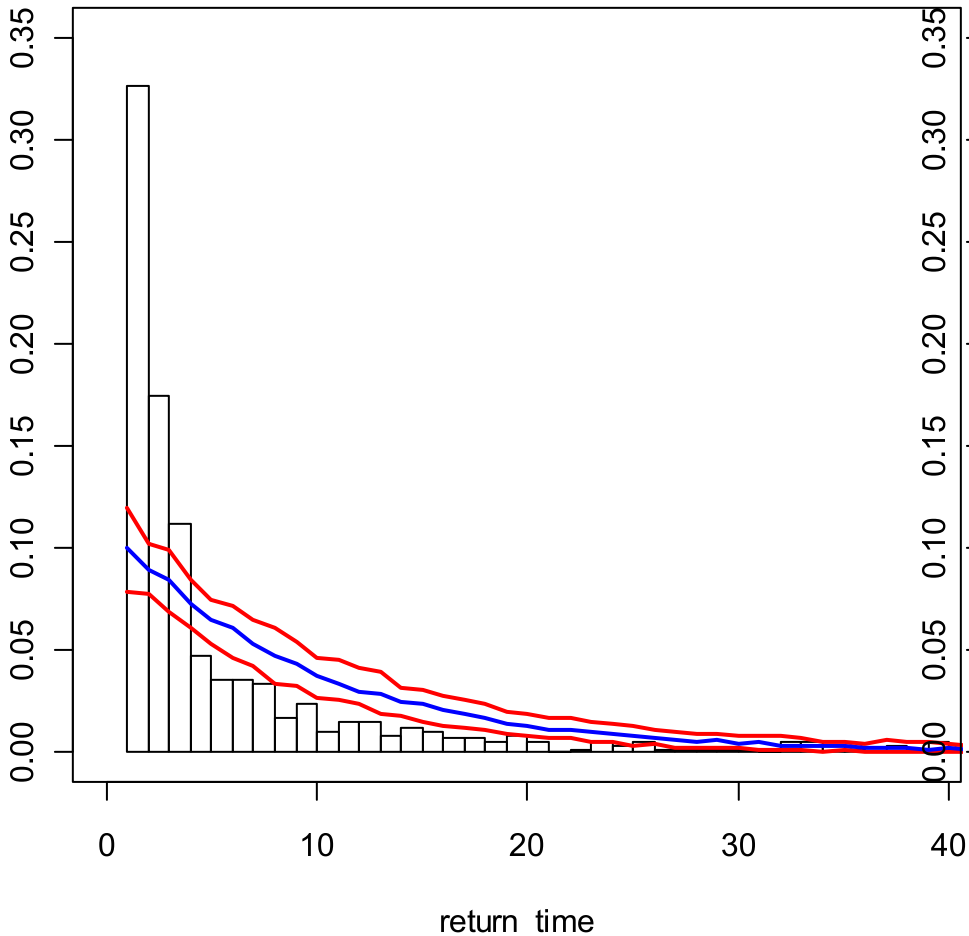
Idea: For v fixed (can do one sided tail), look at the histogram of return times and compare against a geometric distribution.

Example with BAC, v=.05 ⇒ geometric(p=.1)

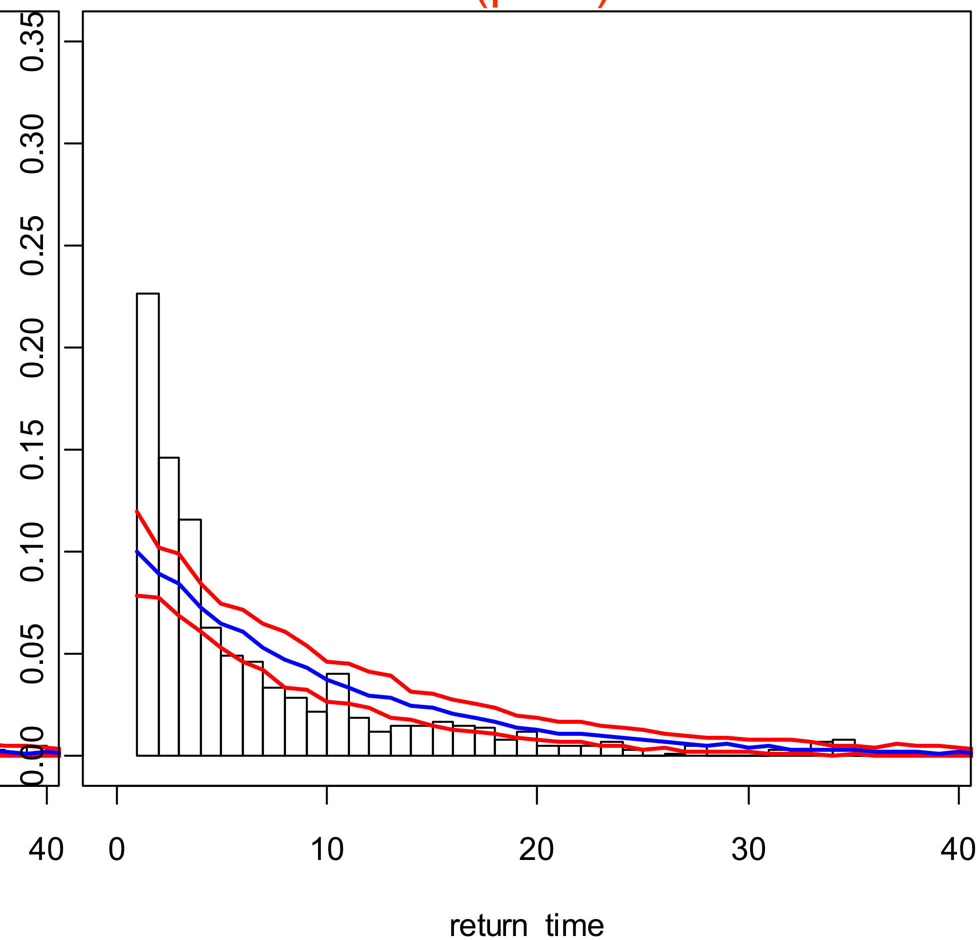


Connections with Return Times (Daily Returns for BAC)

BAC, 2 tail, v=.05 ⇒ G(p=.1)



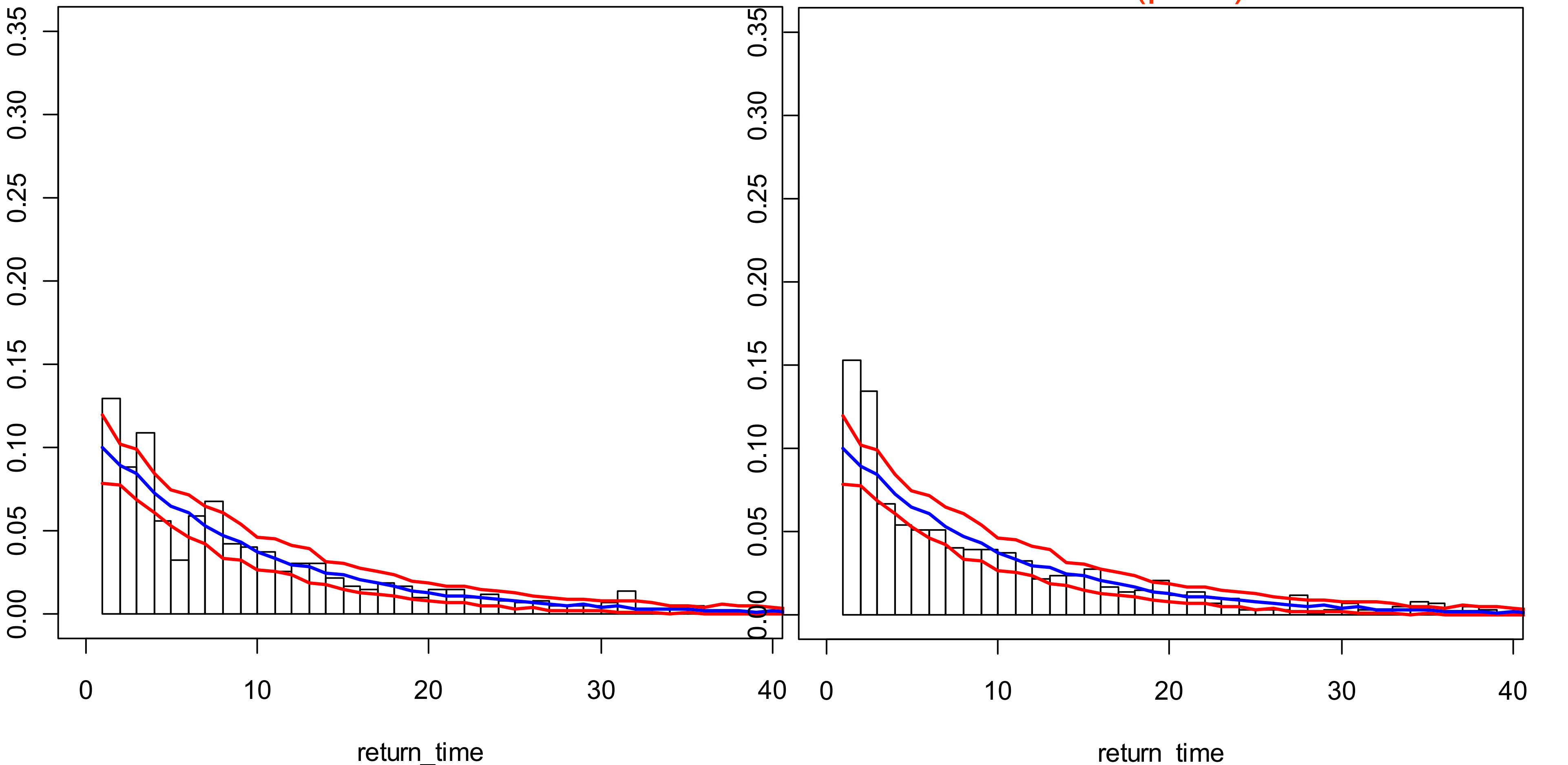
BAC, lower only v=.01 ⇒ G(p=.1):



Connections with Return Times

BAC devolatilized
v=.05 ⇒ G(p=.1)

BAC devolatilized
lower tail only v=.01 ⇒
G(p=.1)



Connections with Return Times (of rare events)

Question: What is the connection with the extremogram?

Answer: The estimated distribution for the return times is exactly the extremogram for specially chosen sets A & B. For example, in the upper tail case, $P(T_1 = 1)$ is estimated by

$$\hat{P}(T=1) = \frac{\sum_{t=1}^{n-1} I_{\{X_t \ge a_m, X_{t+1} \ge a_m\}}}{\sum_{t=1}^{n} I_{\{X_t \ge a_m\}}} = \frac{\#\,\text{consecutive pairs} > a_m}{\#\,\text{observations} > a_m}$$

$$\hat{\rho}_{A,B}(1) = \frac{\frac{m}{n}\sum_{t=1}^{n-1} I_{\{X_t \ge a_m, X_{t+1} \ge a_m\}}}{\frac{m}{n}\sum_{t=1}^{n} I_{\{X_t \ge a_m\}}}$$

Remark: So theory and methodology (permutation/bootstrapping) developed for the extremogram applies to the histogram

Bootstrapping the Extremogram

The stationary bootstrap, introduced by Politis and Romano (1994) seems well suited for the extremogram.

Stationary Bootstrap Setup: Have data $X_1, \ldots, X_n$ and construct BS sample as follows:

- $K_1, K_2, \ldots,$ be iid uniform on $\{1, \ldots, n\}$
- $L_1, L_2, \ldots,$ be iid geometric(p_n)

The BS sample $X_1^*, \ldots, X_n^*$ is given by the first n observations in the sequence.

$$X_{K_1}, \ldots, X_{K_1+L_1-1}, X_{K_2}, \ldots, X_{K_2+L_2-1}, \ldots, X_{K_N}, \ldots, X_{K_N+L_N-1}$$

where

$$N = \inf\{i \geq 1 : L_1 + \cdots + L_i \geq n\}.$$

Bootstrapping the Extremogram

$$X_{K_1},\ldots,X_{K_1+L_1-1},X_{K_2},\ldots,X_{K_2+L_2-1},\ldots,X_{K_N},\ldots,X_{K_N+L_N-1}$$

- $K_1, K_2, \ldots,$ be iid uniform on $\{1,\ldots, n\}$
- $L_1, L_2, \ldots,$ be iid geometric(p_n)

Remarks

- Procedure is similar to the block bootstrap method
- Each block has a random length given by independent geometrics, $L_1, L_2, \ldots$.
- Mean block size is $1/p_n$
- Mean number of blocks is np_n
- By the previous two bullet points, we require

$$p_n \to 0,\ np_n \to \infty.$$

Bootstrapping the Extremogram (cont)

The extremogram, computed from either the sample or BS sample, are ratios of partial sums of the form,

$$\hat{P}_n(C) = \frac{m_n}{n}\sum_{t=1}^{n} I_{\{a_m^{-1}X_t \in C\}} \quad \text{and} \quad \hat{P}_n^*(C) = \frac{m_n}{n}\sum_{t=1}^{n} I_{\{a_m^{-1}X_t^* \in C\}}.$$

Theorem . Assuming our general setup (mixing conditions + regular variation, etc), and the growth conditions,

$$np_n \to \infty, \quad np^2/m_n \to \infty,$$

we have $E^*\hat{P}_n^*(C) \xrightarrow{P} \mu(C)$ and $ms_n^2 = \mathrm{Var}^*((n/m)^{1/2}\hat{P}_n^*(C)) \xrightarrow{P} \sigma^2(C).$

Moreover,

$$\sup_x |P((n/m)^{1/2}(ms_n^2)^{-1/2}(\hat{P}_n^*(C) - \hat{P}_n(C)) \le x \mid X_1,\ldots,X_n) - \Phi(x)| \xrightarrow{P} 0$$

Bootstrapping the Extremogram (cont)

The sample extremogram and its BS counterpart are:

$$\hat{\rho}_{A,B}(h) = \frac{\frac{m}{n}\sum_{t=1}^{n-h} I_{\{a_m^{-1}X_t \in A, a_m^{-1}X_{t+h} \in B\}}}{\frac{m}{n}\sum_{t=1}^{n} I_{\{a_m^{-1}X_t \in A\}}} \qquad \hat{\rho}^*_{A,B}(h) = \frac{\frac{m}{n}\sum_{t=1}^{n-h} I_{\{a_m^{-1}X_t^* \in A, a_m^{-1}X_{t+h}^* \in B\}}}{\frac{m}{n}\sum_{t=1}^{n} I_{\{a_m^{-1}X_t^* \in A\}}}$$

Theorem . Assuming our general setup (mixing conditions + regular variation, etc), and the growth conditions,

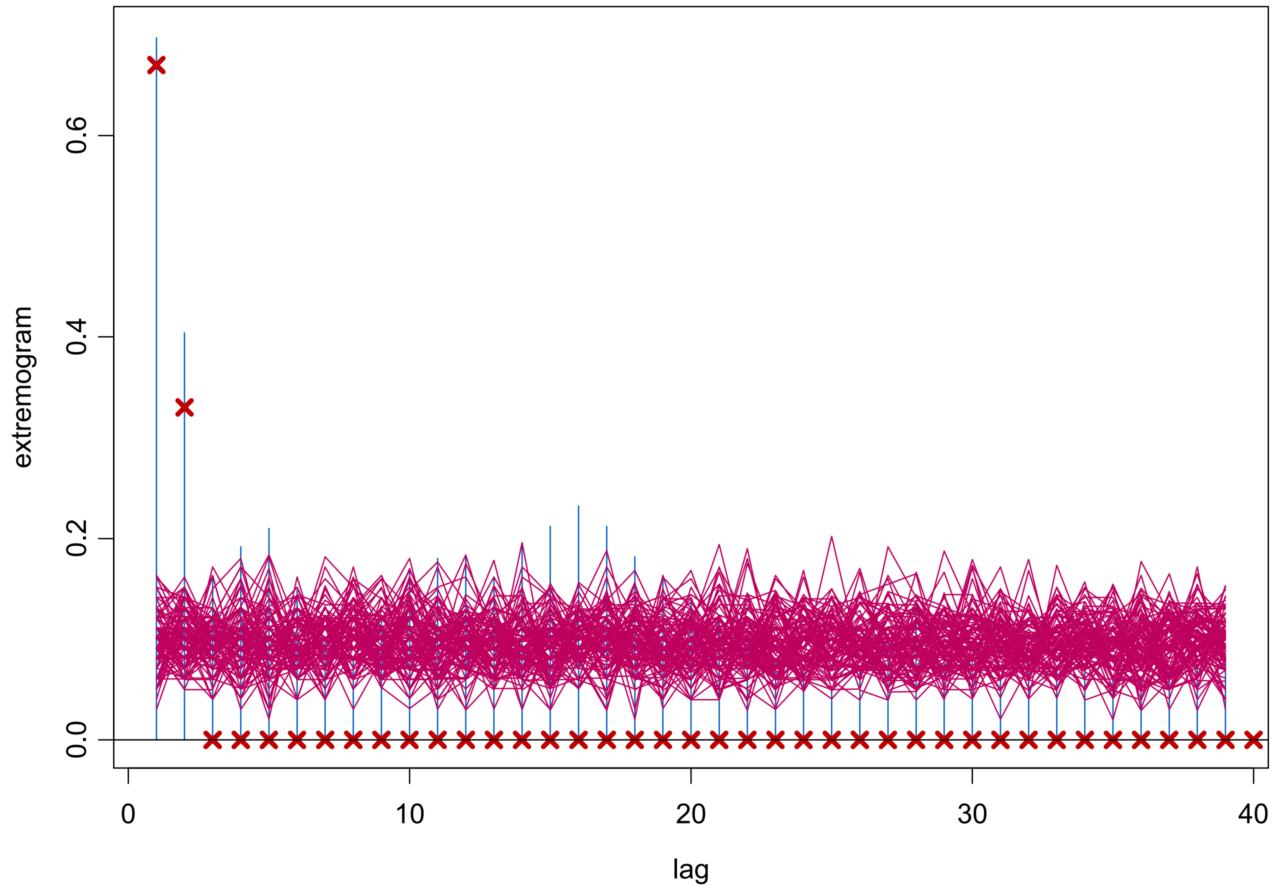
$$np_n \to \infty, \quad np^2/m_n \to \infty,$$

we have

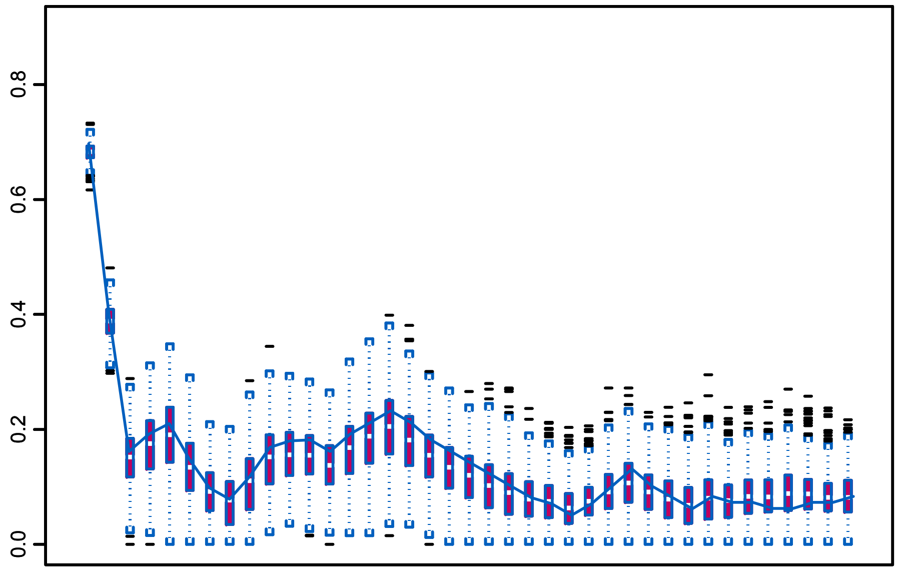
$$\sup_x |P((n/m)^{1/2}(\hat{\rho}^*_{A,B}(h) - \hat{\rho}_{A,B}(h)) \le x \,|\, X_1,\ldots,X_n) -$$

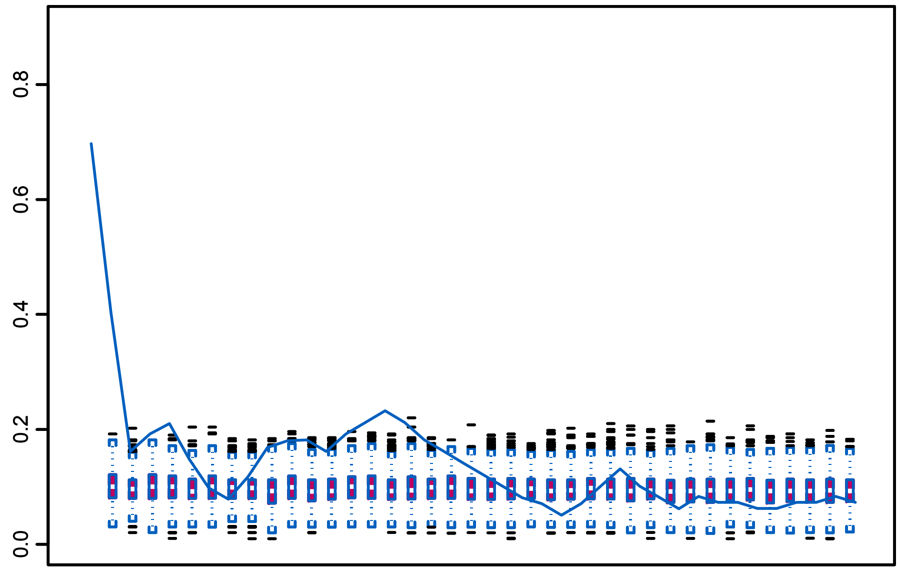
$$P((n/m)^{1/2}(\hat{\rho}_{A,B}(h) - \rho_m(h)) \le x)| \xrightarrow{P} 0$$

Bootstrap of the Extremogram of the Max-MA(2)



Bootstrap of the Extremogram of the Max-MA(2)





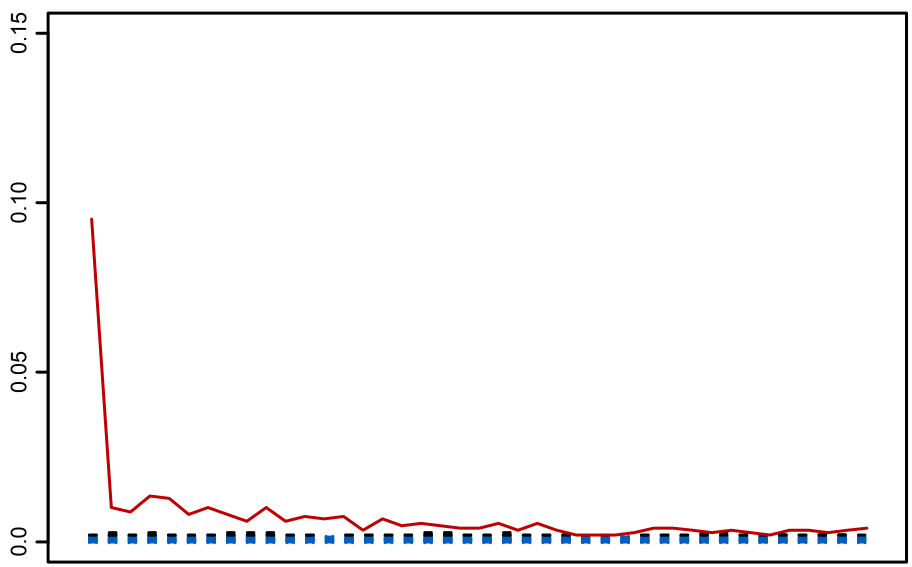
p_n =.02 (mean block size is 50) p_n = 1 (mean block size is 1)

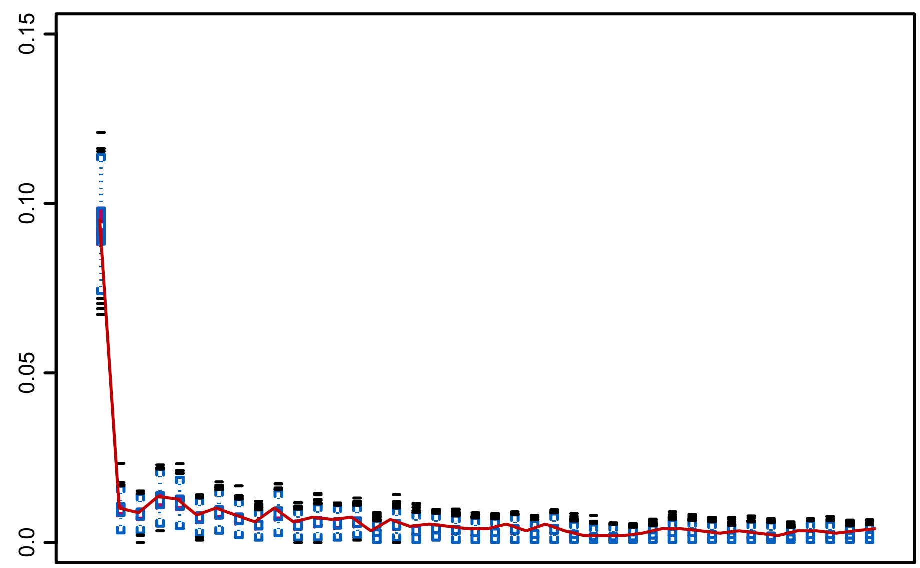
BS reps = 1000

Extremogram of a GARCH(1,1)

GARCH(1,1): $X_t = (.1 + .14\, X^2_{t-1} + .83\sigma^2_{t-1})^{1/2} Z_t$, $\{Z_t\} \sim$ IID N(0,1), $n=10^6$

3-dim extremogram ($\lim_n P(\min(X_h, X_{h+1}) > n^{1/\alpha} \mid X_0 > n^{1/\alpha})$)





$p_n = 1$ (mean block size is 1)

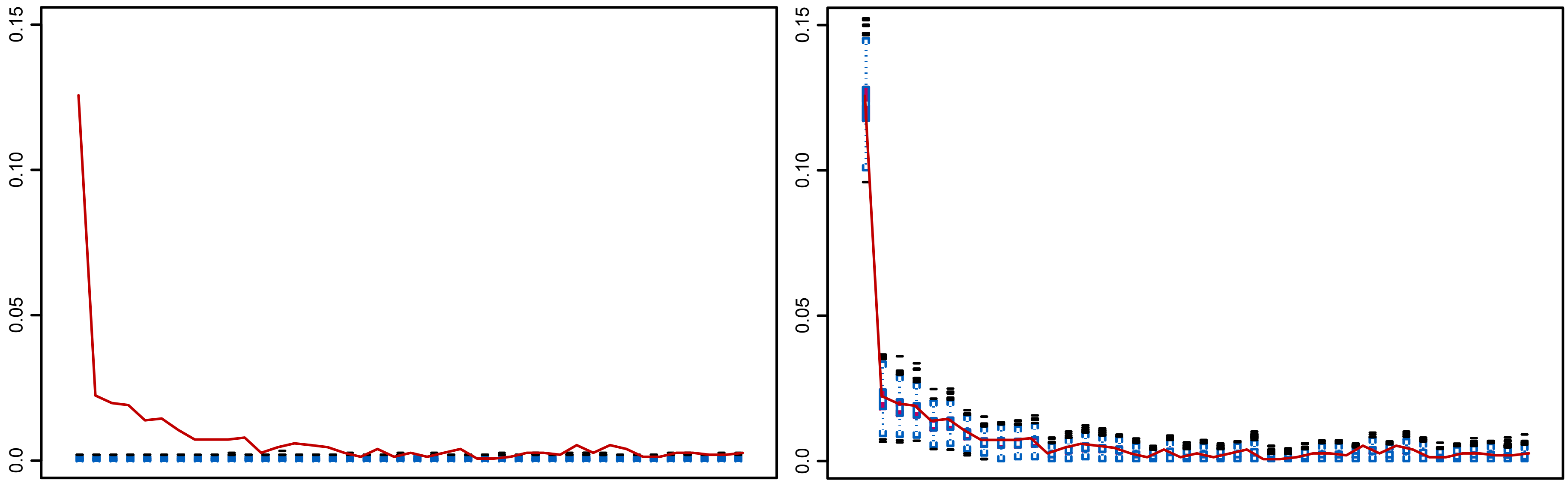
$p_n = .02$ (mean block size is 50)

BS reps = 1000

Extremogram of a SV

SV Process: $X_t = \sigma_t Z_t$, $\{Z_t\} \sim$ IID t_4 ; $\log \sigma_t = .9 \log \sigma_{t-1} + \varepsilon_t$, $n=10^6$

3-dim extremogram ($\lim_n P(\min(X_h, X_{h+1}) > n^{1/\alpha} \mid X_0 > n^{1/\alpha}$))



$p_n = 1$ (mean block size is 1)

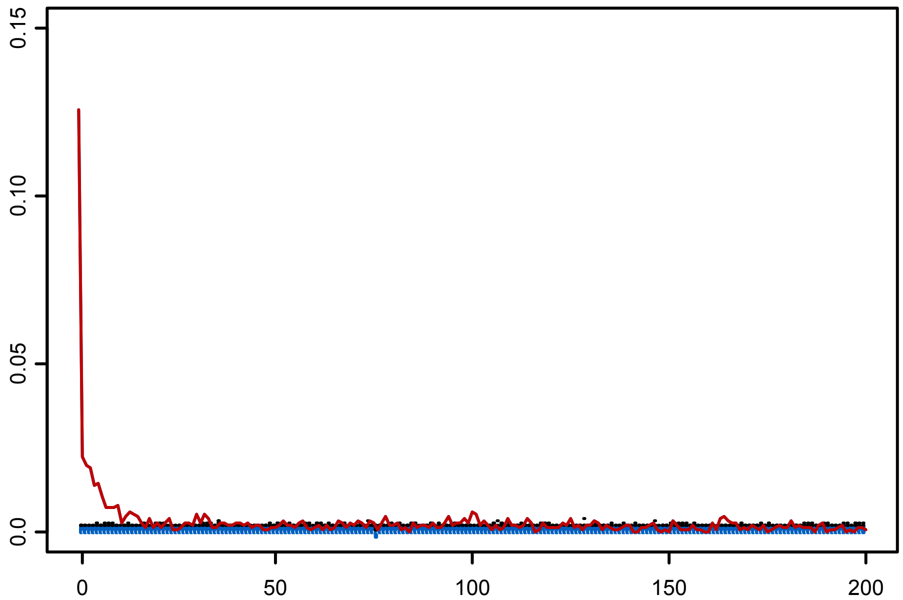
$p_n = .02$ (mean block size is 50)

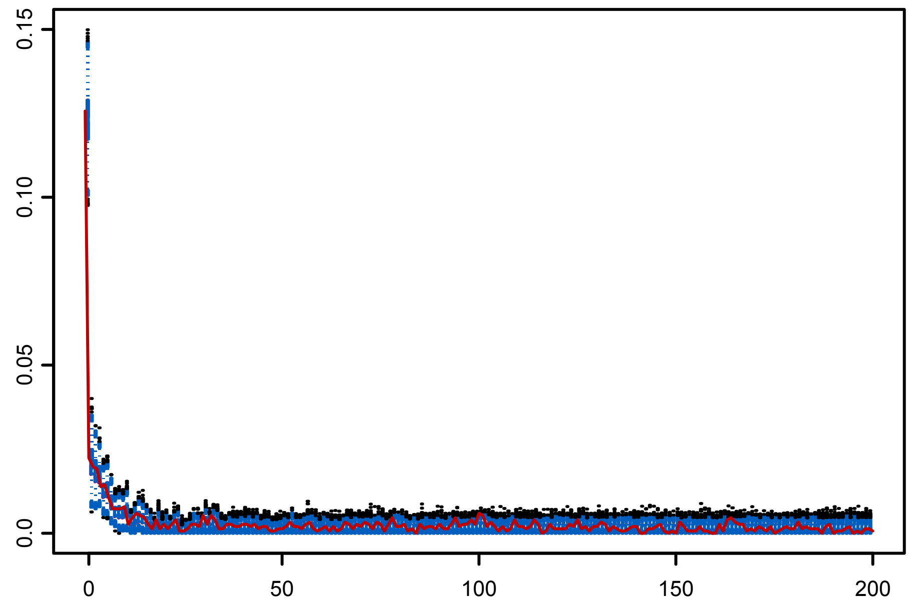
BS reps = 1000

Extremogram of a SV

SV Process: $X_t=\sigma_t Z_t$, $\{Z_t\} \sim$ IID t_4 ; $\log \sigma_t = .9 \log \sigma_{t-1} + \varepsilon_t$, $n=10^6$

3-dim extremogram ($\lim_n P(\min(X_h,X_{h+1}) > n^{1/\alpha} \mid X_0 > n^{1/\alpha}$))



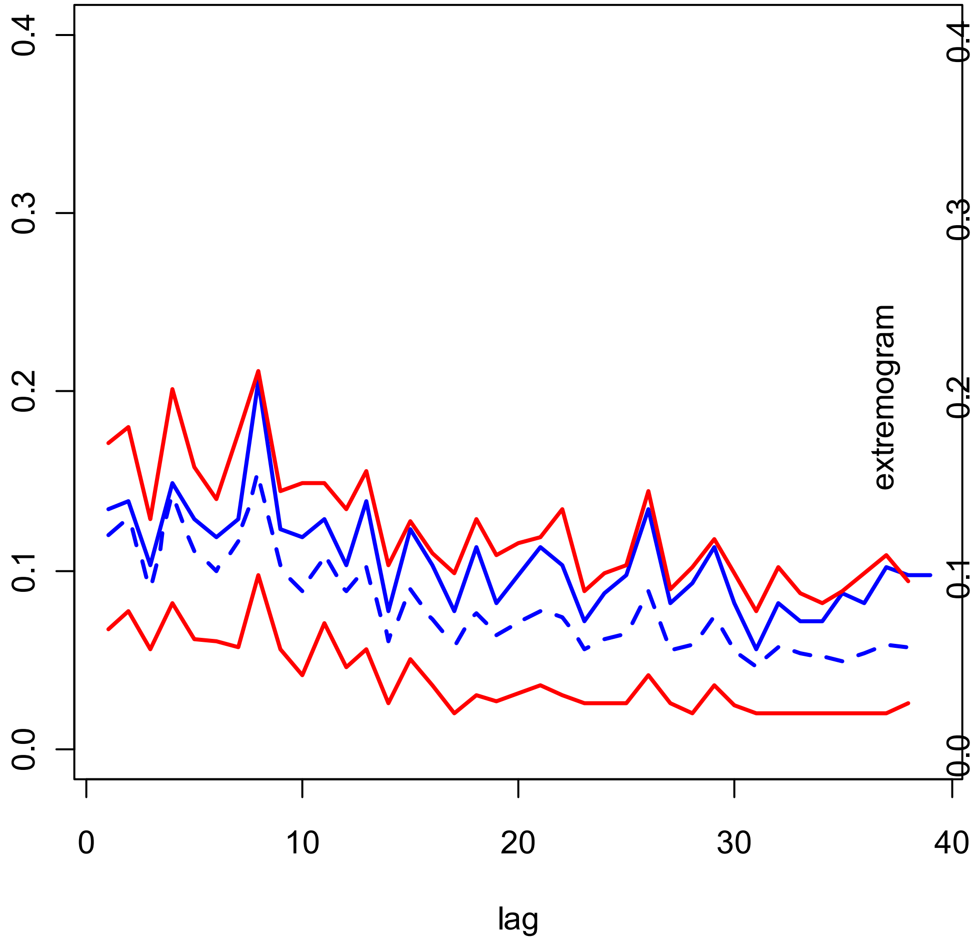


p_n =1 (mean block size is 1)

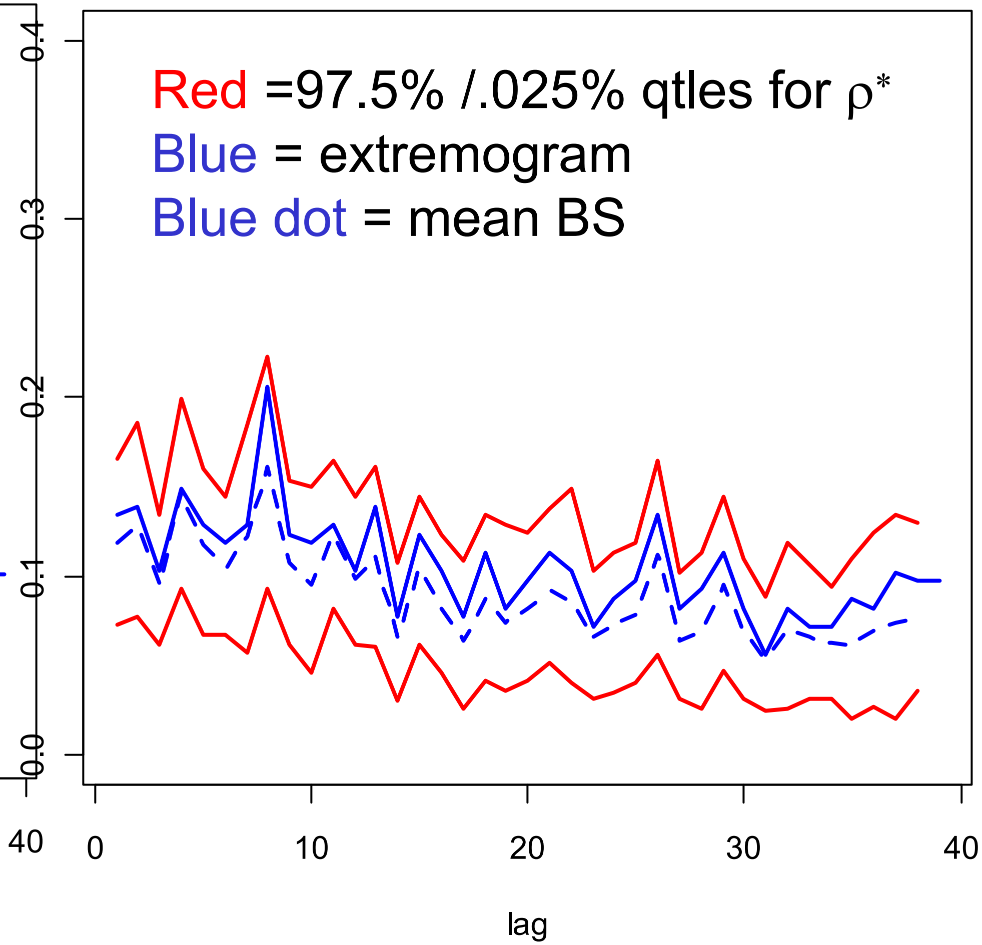
p_n = .02 (mean block size is 50)

BS reps = 1000

Application to FTSE (lower tail)

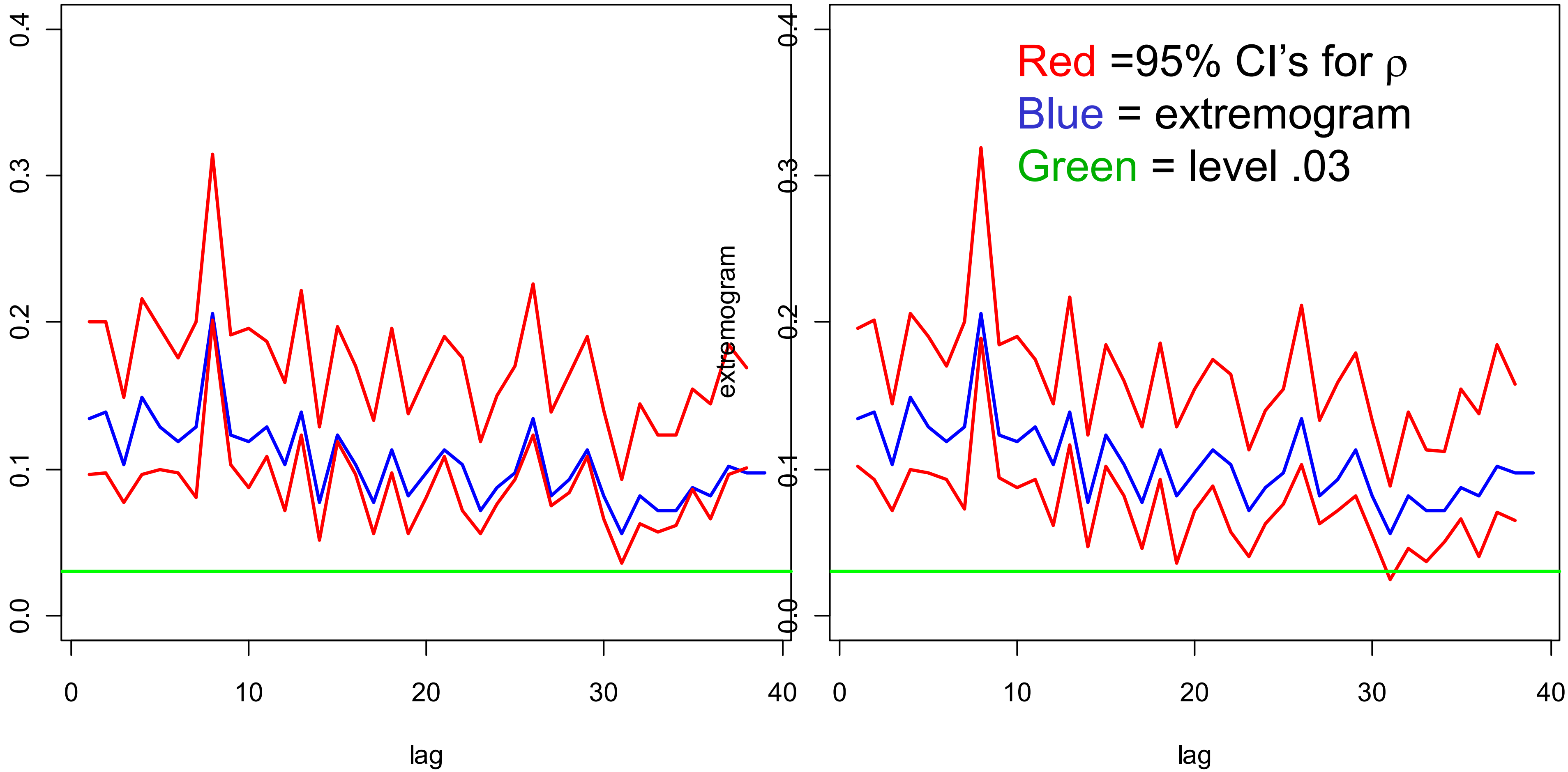


p_n =1/50



p_n =1/200

Application to FTSE (lower tail)



p_n =1/50 p_n =1/200

References

Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (Eds.) (2009) The Handbook of Financial Time Series. Springer, Heidelberg.

Basrak, B., Davis, R.A. and Mikosch, T. (2002) Regular variation of GARCH processes. Stoch. Proc. Appl. 99, 95–116.

Basrak, B. and Segers J. (2009) Regularly varying multivariate time series. Stoch. Proc. Appl. 119, 1055–1080.

Billingsley, P. (1968) Convergence of Probability Measures. (1968) Wiley, New York.

Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) Regular Variation. Cambridge University Press, Cambridge.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31, 307–327.

Bougerol, P. and Picard, N. (1992) Strict stationarity of generalized autoregressive processes. Ann. Probab. 20, 1714–1730.

Breiman, L. (1965) On some limit theorems similar to the arc-sin law. Theory Probab. Appl. 10, 323–331.

Brockwell, P.J. and Davis, R.A. (1991) Time Series: Theory and Methods, 2nd edition Springer-Verlag, New York.

Davis, R.A. and Hsing, T. (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. Ann. Probab. 23, 879–917.

Davis, R.A. and Mikosch, T. (1998) Limit theory for the sample ACF of stationary process with heavy tails with applications to ARCH. Ann. Statist. 26, 2049–2080.

Davis, R.A. and Mikosch, T. (2009) The extremogram: a correlogram for extreme events. Bernoulli 15, 977–1009.

Davis, R.A. and Mikosch, T. (2009) Probabilistic models of stochastic volatility models. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (Eds.) The Handbook of Financial Time Series. Springer, Heidelberg, pp. 255–268.

Davis, R.A. and Mikosch, T. (2009) Extreme value theory for GARCH processes. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (Eds.) The Handbook of Financial Time Series. Springer, Heidelberg, pp. 187–200.

Davis, R.A., Mikosch, T. and Cribben, I. (2012) Towards estimating extremal serial dependence via the bootstrapped extremogram. J. Econometrics 170, 142–152.

Davis, R.A. and Resnick, S.I. (1985) Limit theory for

Davis, R.A. and Resnick, S.I. (1996) Limit theory for bilinear processes with heavy-tailed noise Ann. Appl. Probab. 6, 1191–1210.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.

Engle, R.F. and Bollerslev, T. (1986) Modelling the persistence of conditional variances. With comments and a reply by the authors. Econometric Rev. 5, 1–87.

Feller, W. (1971) An Introduction to Probability Theory and Its Applications. Vol. II. Second edition. Wiley, New York.

Francq, C. and Zakoian, J.-M. (2010) GARCH Models. Wiley, Chichester.

Goldie, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. Ann. Appl. Probab. 1, 126–166.

Goldie, C.M. and Grübel, R. (1996) Perpetuities with thin tails. Adv. Appl. Probab. 28 , 463–480.

Haan, L. de, Resnick, S.I., Rootzén, H. and Vries, C.G. de (1989) Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. Stoch. Proc. Appl. 32, 213–224.

Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. Acta Math. 131, 207–248.

Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. Acta Math. 131, 207–248.

Konstantinides, D. and Mikosch, T. (2005) Stochastic recurrence equations with heavy-tailed innovations. Ann. Probab. 33, 1992–2035.

Leadbetter, M.R. (1983) Extremes and local dependence of stationary sequences. Z. Wahrscheinlichkeitstheorie verw. Gebiete 65, 291–306.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) Extremes and Related Properties of Random Sequences and Processes. Springer, Berlin.

Mikosch, T. and Stărică, C. (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. Ann. Statist. 28, 1427–1451.

Resnick, S.I. (1986) Point processes, regular variation and weak convergence. Adv. Appl. Prob. 18, 66–138.

Resnick, S.I. (1987) Extreme Values, Regular Variation, and Point Processes. Springer, New York.

Resnick, S.I. (2007) Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. Springer, New York.

Rootzén, H. (1978) Extremes of moving averages of stable processes. Ann. Probab. 6, 847–869.

Rootzén, H. (1986) Extreme value theory for moving average processes. Ann. Probab. 14, 612–652.

Rvaceva, E.L. (1962) On domains of attraction of multi-dimensional distributions. Select. Transl. Math. Statist. and Probability of the AMS 2. 183–205.

Samorodnitsky, G. and Taqqu, M.S. (1994) Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance. Chapman and Hall, London.

Segers, J. (2005) Approximate distributions of clusters of extremes. Stat. Probab. Letters 74, 330–336.

Shephard, N. and Andersen, T.G. (2009) Stochastic volatility: origins and overview. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (Eds.) The Handbook of Financial Time Series. Springer, Heidelberg, pp. 233–254.

Thank You