Threshold Selection for Multivariate Heavy-Tailed Data

Phyllis Wan[*,**], Richard Davis[*]

Columbia University[*], Erasmus University Rotterdam[**]

June 21, 2018

Regular variation

- Univariate regularly varying: $X \in \mathbb{R}_+$, $X \sim RV(\alpha)$ if

$$\lim_{t \to \infty} \mathbb{P}[X > tx | X > t] = c(x), \quad x \geq 1.$$

  - c is of the form $c(x) = x^{-\alpha}$, $\alpha > 0$.

Regular variation

- Univariate regularly varying: $X \in \mathbb{R}_+$, $X \sim RV(\alpha)$ if

$$\lim_{t\to\infty} \mathbb{P}[X > tx | X > t] = c(x), \quad x \geq 1.$$

  - c is of the form $c(x) = x^{-\alpha}$, $\alpha > 0$.

- Multivariate regularly varying: $\mathbf{X} \in \mathbb{R}_+^d$, $\mathbf{X} \sim \mathbf{MRV}(\alpha)$ if

$$\lim_{t\to\infty} \mathbb{P}[\mathbf{X} > t\mathbf{x} | \mathbf{X} > t\mathbf{1}] = \nu(\mathbf{x}), \quad \mathbf{x} \geq \mathbf{1}.$$

  - ν satisfies $\nu(s\mathbf{x}) = s^{-\alpha}\nu(\mathbf{x})$, $\alpha > 0$

Regular variation

- Univariate regularly varying: $X \in \mathbb{R}_+$, $X \sim RV(\alpha)$ if

$$\lim_{t \to \infty} \mathbb{P}[X > tx | X > t] = c(x), \quad x \geq 1.$$

  - c is of the form $c(x) = x^{-\alpha}$, $\alpha > 0$.

- Multivariate regularly varying: $\mathbf{X} \in \mathbb{R}_+^d$, $\mathbf{X} \sim \mathbf{MRV}(\alpha)$ if

$$\lim_{t \to \infty} \mathbb{P}[\mathbf{X} > t\mathbf{x} | \mathbf{X} > t\mathbf{1}] = \nu(\mathbf{x}), \quad \mathbf{x} \geq \mathbf{1}.$$

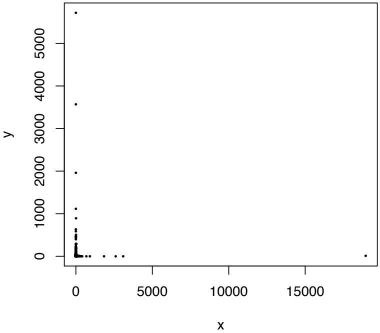
  - ν satisfies $\nu(s\mathbf{x}) = s^{-\alpha}\nu(\mathbf{x})$, $\alpha > 0$

- Let $(R, \boldsymbol{\Theta}) = (\|\mathbf{X}\|, \frac{\mathbf{X}}{\|\mathbf{X}\|})$, then $\mathbf{X} \sim \mathbf{MRV}(\alpha)$ if and only if

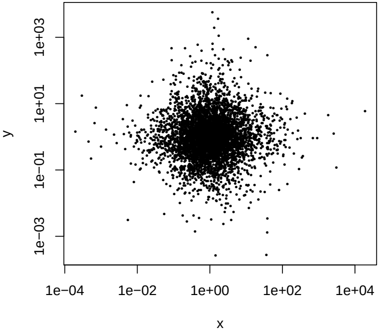
  1. $R \sim$ Univariate $RV(\alpha)$
  2. $P(\boldsymbol{\Theta} \in \cdot | R > r) \to S(\cdot), \quad r \to \infty.$

  - In other words, $\boldsymbol{\Theta}$ becomes independent of R as $R \to \infty$.
  - S characterizes the extremal dependence.

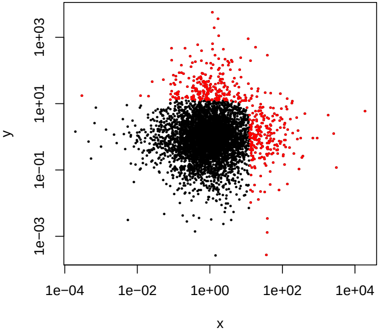
Example: $X_i, Y_i \overset{iid}{\sim} |t_1|$

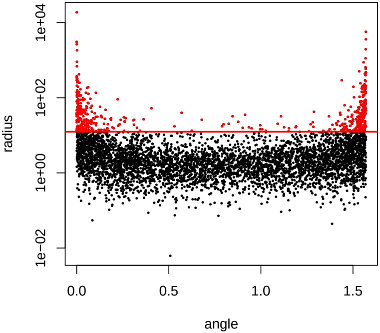


Example: $X_i, Y_i \stackrel{iid}{\sim} |t_1|$

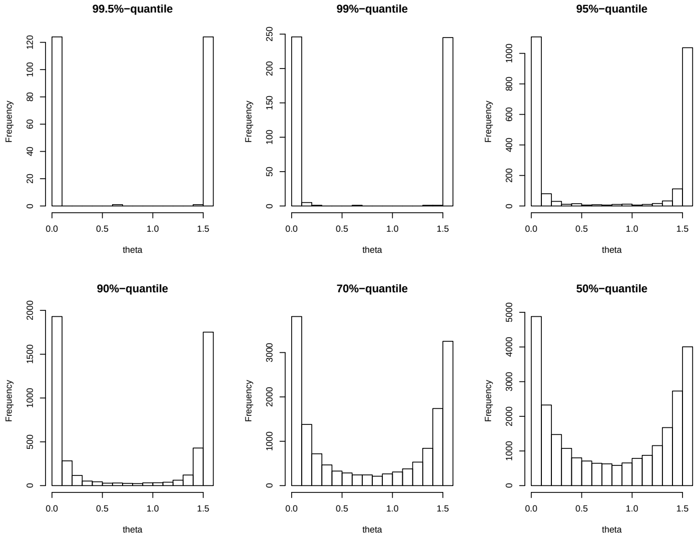


Example: $X_i, Y_i \overset{iid}{\sim} |t_1|$



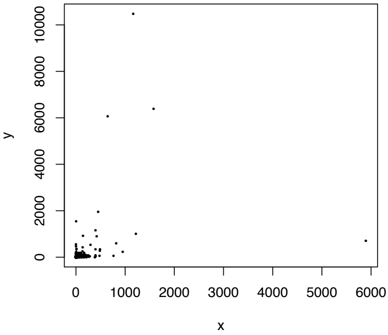


Example: $X_i, Y_i \stackrel{iid}{\sim} |t_1|$

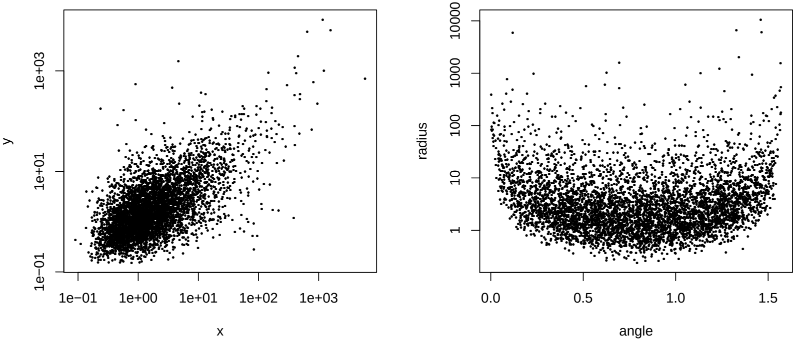


Example: $(X_i, Y_i) \overset{iid}{\sim}$ *Bilogistic*

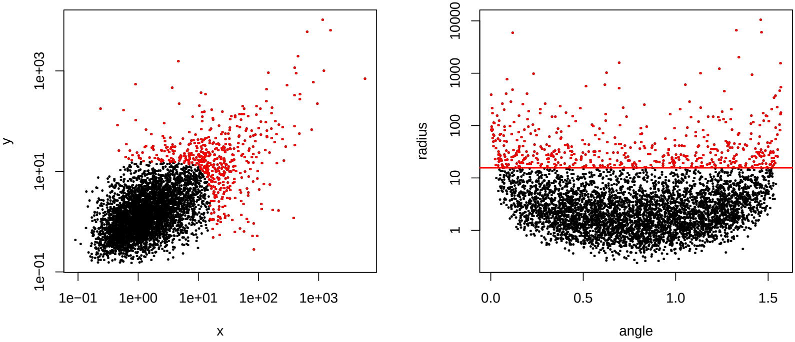
- $F(x, y) = \exp\left\{-(x^{-1/s} + y^{-1/s})^s\right\}$
- $s = 0.6$



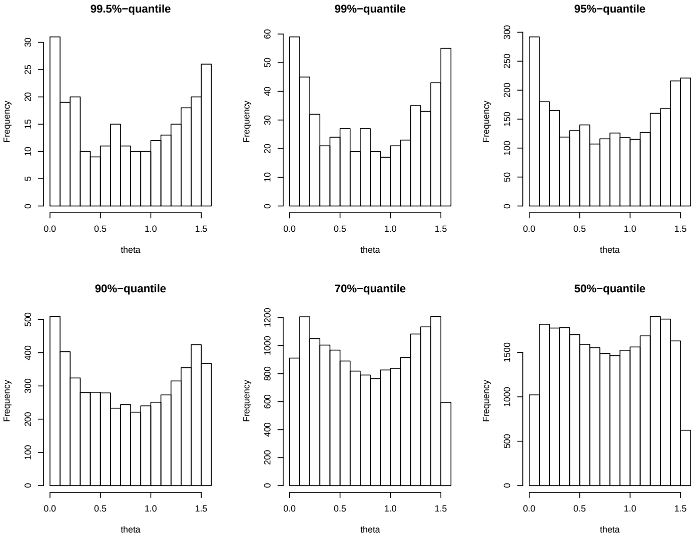
Example: $(X_i, Y_i) \overset{iid}{\sim}$ *Bilogistic*



Example: $(X_i, Y_i) \overset{iid}{\sim}$ *Bilogistic*



Example: $(X_i, Y_i) \overset{iid}{\sim}$ *Bilogistic*



Estimating $S(\cdot)$, the limiting angular distribution

Observe $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathbf{MRV}(\alpha)$ and $(R_i, \mathbf{\Theta}_i) = (\|\mathbf{X}_i\|, \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|})$. We know

$$P(\mathbf{\Theta} \in \cdot | R > r) \to S(\cdot), \quad r \to \infty.$$

How to estimate $S(\cdot)$?

- Look at the subset $\mathbf{\Theta}_{i_1}, \ldots, \mathbf{\Theta}_{i_K}$ where $R_{i_k} > r_0$ for r_0 large

Estimating $S(\cdot)$, the limiting angular distribution

Observe $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathbf{MRV}(\alpha)$ and $(R_i, \mathbf{\Theta}_i) = (\|\mathbf{X}_i\|, \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|})$. We know

$$P(\mathbf{\Theta} \in \cdot | R > r) \to S(\cdot), \quad r \to \infty.$$

How to estimate $S(\cdot)$?

- Look at the subset $\mathbf{\Theta}_{i_1}, \ldots, \mathbf{\Theta}_{i_K}$ where $R_{i_k} > r_0$ for r_0 large

How large should r_0 be?

- Conditional on $R > r_0$, $(R, \mathbf{\Theta})$ are approximately independent

Estimating $S(\cdot)$, the limiting angular distribution

Observe $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathbf{MRV}(\alpha)$ and $(R_i, \boldsymbol{\Theta}_i) = (\|\mathbf{X}_i\|, \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|})$. We know

$$P(\boldsymbol{\Theta} \in \cdot | R > r) \to S(\cdot), \quad r \to \infty.$$

How to estimate $S(\cdot)$?

- Look at the subset $\boldsymbol{\Theta}_{i_1}, \ldots, \boldsymbol{\Theta}_{i_K}$ where $R_{i_k} > r_0$ for r_0 large

How large should r_0 be?

- Conditional on $R > r_0$, $(R, \boldsymbol{\Theta})$ are approximately independent

How to measure the dependence between R and $\boldsymbol{\Theta}$?

- R is heavy-tailed – may not even have 1st moment!
- $\boldsymbol{\Theta}$ could be multi-dimensional
- Solution: **distance covariance**

Distance covariance

- Feuerverger (1993), Székely et al. (2007), Meintanis & Iliopoulos (2008).
- $X \in \mathbb{R}^p$, $Y \in \mathbb{R}^q$, let φ denote the characteristic function, then

$$X \perp Y \iff \varphi_{X,Y} = \varphi_X \varphi_Y.$$

- *Distance covariance* w.r.t. weight measure $\mu(s,t)$

$$T(X,Y;\mu) = \int_{\mathbb{R}^{p+q}} |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \mu(ds,dt).$$

- *Distance correlation*

$$R(X,Y;\mu) = \frac{T(X,Y;\mu)}{\sqrt{T(X,X;\mu)T(Y,Y;\mu)}} \in [0,1].$$

Distance covariance

$$T(X, Y; \mu) = \int_{\mathbb{R}^{p+q}} |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \mu(ds, dt)$$

Empirical version?

Distance covariance

$$T(X, Y; \mu) = \int |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \mu(ds, dt)$$

$$= \int |\mathbb{E}e^{isX+itY} - \mathbb{E}e^{isX}\mathbb{E}e^{itY}|^2 \mu(ds, dt)$$

Distance covariance

$$\begin{aligned} T(X, Y; \mu) &= \int |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \mu(ds, dt) \\ &= \int |\mathbb{E}e^{isX+itY} - \mathbb{E}e^{isX}\mathbb{E}e^{itY}|^2 \mu(ds, dt) \end{aligned}$$

$\Big($Let X', Y', Y'', Y''' be independent copies of $X, Y\Big)$

Distance covariance

$$
\begin{aligned}
T(X, Y; \mu) &= \int |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \mu(ds, dt) \\
&= \int |\mathbb{E}e^{isX+itY} - \mathbb{E}e^{isX}\mathbb{E}e^{itY}|^2 \mu(ds, dt)
\end{aligned}
$$

(Let X', Y', Y'', Y''' be independent copies of X, Y)

$$
\begin{aligned}
&= \int \Big(\mathbb{E}e^{is(X-X')+it(Y-Y')} + \mathbb{E}e^{is(X-X')}e^{it(Y''-Y''')} \\
&\qquad -\mathbb{E}e^{is(X-X')+it(Y-Y'')} - \mathbb{E}e^{-is(X-X')-it(Y-Y'')}\Big)\, \mu(ds, dt)
\end{aligned}
$$

Distance covariance

$$T(X, Y; \mu) = \int |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \mu(ds, dt)$$

$$= \int |\mathbb{E}e^{isX+itY} - \mathbb{E}e^{isX}\mathbb{E}e^{itY}|^2 \mu(ds, dt)$$

$$\left(\text{Let } X', Y', Y'', Y''' \text{ be independent copies of } X, Y\right)$$

$$= \int \Big(\mathbb{E}e^{is(X-X')+it(Y-Y')} + \mathbb{E}e^{is(X-X')}e^{it(Y''-Y''')}$$

$$-\mathbb{E}e^{is(X-X')+it(Y-Y'')} - \mathbb{E}e^{-is(X-X')-it(Y-Y'')}\Big)\,\mu(ds, dt)$$

$$\left(\text{Let } h(x,y) = \mathrm{Re}\left(\int e^{isx+ity}\mu(ds, dt)\right)\right)$$

Distance covariance

$$\begin{aligned} T(X, Y; \mu) &= \int |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \mu(ds, dt) \\ &= \int |\mathbb{E}e^{isX+itY} - \mathbb{E}e^{isX}\mathbb{E}e^{itY}|^2 \mu(ds, dt) \end{aligned}$$

$$\left(\text{Let } X', Y', Y'', Y''' \text{ be independent copies of } X, Y\right)$$

$$\begin{aligned} &= \int \Big(\mathbb{E}e^{is(X-X')+it(Y-Y')} + \mathbb{E}e^{is(X-X')}e^{it(Y''-Y''')} \\ &\qquad -\mathbb{E}e^{is(X-X')+it(Y-Y'')} - \mathbb{E}e^{-is(X-X')-it(Y-Y'')}\Big)\,\mu(ds, dt) \end{aligned}$$

$$\left(\text{Let } h(x, y) = \text{Re}\left(\int e^{isx+ity}\mu(ds, dt)\right)\right)$$

$$\begin{aligned} &= \mathbb{E}h(X - X', Y - Y') + \mathbb{E}h(X - X', Y'' - Y''') \\ &\quad -2\mathbb{E}h(X - X', Y - Y'') \end{aligned}$$

Distance covariance

$$T(X, Y; \mu) = \int_{\mathbb{R}^{p+q}} \left|\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)\right|^2 \mu(ds, dt)$$

Empirical version

$$\begin{aligned} T_n(X, Y; \mu) \quad = \quad & \frac{1}{n^2} \sum_{j,k=1}^{n} h(X_j - X_k, Y_j - Y_k) \\ & + \frac{1}{n^4} \sum_{j,k,l,r=1}^{n} h(X_j - X_k, Y_l - Y_r) \\ & - \frac{2}{n^3} \sum_{k,l,r=1}^{n} h(X_j - X_k, Y_j - Y_l) \end{aligned}$$

Distance covariance

$$T(X, Y; \mu) = \int_{\mathbb{R}^{p+q}} |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \mu(ds, dt)$$

Choice of μ?

- Székely et al. (2007): $\mu(s,t) \propto |s|_q^{-\alpha-q}|t|_p^{-\alpha-p} ds\, dt$, for $0 < \alpha < 2$.
  - $h(x - x', y - y') = |x - x'|_p^\alpha |y - y'|_q^\alpha$
  - Requires $E|X|_p^\alpha + E|Y|_q^\alpha + E|X|_p^\alpha |Y|_q^\alpha < \infty$

Distance covariance

$$T(X, Y; \mu) = \int_{\mathbb{R}^{p+q}} |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \mu(ds, dt)$$

Choice of μ?

- Székely et al. (2007): $\mu(s,t) \propto |s|_q^{-\alpha-q}|t|_p^{-\alpha-p} ds\, dt$, for $0 < \alpha < 2$.
  - $h(x - x', y - y') = |x - x'|_p^\alpha |y - y'|_q^\alpha$
  - Requires $E|X|_p^\alpha + E|Y|_q^\alpha + E|X|_p^\alpha |Y|_q^\alpha < \infty$

- $\mu(ds, dt) = \mu_S(ds)\mu_T(dt)$, product of probability measures
  - $h(x - x', y - y') = \varphi_S(x - x')\varphi_T(y - y')$.
  - No constraints on X, Y
  - E.g. Normal, $h(x - x') = \exp(-\frac{\sigma^2}{2}|x - x'|^2)$
  - E.g. Cauchy, $h(x - x') = \exp(-\gamma|x - x'|)$

Limit theory of distance covariance (Davis et al., 2018)

Consistency

Let $\{(X_t, Y_t)\}$ be stationary and ergodic, then

$$T_n(X, Y; \mu) \stackrel{a.s.}{\to} T(X, Y; \mu).$$

Limiting distribution

Further let $\{(X_t, Y_t)\}$ be α-mixing with $\sum_{h=1}^{\infty} \alpha_h^{1/r} < \infty$, $1 < r < 2$.

- If $\{X_t\}$ and $\{Y_t\}$ are independent, then

$$nT_n(X, Y; \mu) \stackrel{d}{\to} \int |Q_{X,Y}|^2 d\mu.$$

  where $Q_{X,Y}$ is a centered Gaussian process.

- If $\{X_t\}$ and $\{Y_t\}$ are dependent, then

$$\sqrt{n}\,(T_n(X, Y; \mu) - T(X, Y; \mu)) \stackrel{d}{\to} \int Q'_{X,Y} d\mu.$$

  where $Q'_{X,Y}$ is a centered Gaussian process.

- **Test of independence.**

Limit theory of distance covariance (Davis et al., 2018)

Consistency

Let $\{(X_t, Y_t)\}$ be stationary and ergodic, then

$$T_n(X, Y; \mu) \stackrel{a.s.}{\to} T(X, Y; \mu).$$

Limiting distribution

Further let $\{(X_t, Y_t)\}$ be α-mixing with $\sum_{h=1}^{\infty} \alpha_h^{1/r} < \infty$, $1 < r < 2$.

- If $\{X_t\}$ and $\{Y_t\}$ are independent, then

$$nT_n(X, Y; \mu) \stackrel{d}{\to} \int |Q_{X,Y}|^2 d\mu.$$

  where $Q_{X,Y}$ is a centered Gaussian process.

- If $\{X_t\}$ and $\{Y_t\}$ are dependent, then

$$\sqrt{n}\,(T_n(X, Y; \mu) - T(X, Y; \mu)) \stackrel{d}{\to} \int Q'_{X,Y} d\mu.$$

  where $Q'_{X,Y}$ is a centered Gaussian process.

- **Test of independence.**

Limit theory of distance covariance (Davis et al., 2018)

Consistency

Let $\{(X_t, Y_t)\}$ be stationary and ergodic, then

$$T_n(X, Y; \mu) \stackrel{a.s.}{\to} T(X, Y; \mu).$$

Limiting distribution

Further let $\{(X_t, Y_t)\}$ be α-mixing with $\sum_{h=1}^{\infty} \alpha_h^{1/r} < \infty$, $1 < r < 2$.

- If $\{X_t\}$ and $\{Y_t\}$ are independent, then

$$nT_n(X, Y; \mu) \stackrel{d}{\to} \int |Q_{X,Y}|^2 d\mu.$$

  where $Q_{X,Y}$ is a centered Gaussian process.

- If $\{X_t\}$ and $\{Y_t\}$ are dependent, then

$$\sqrt{n}\,(T_n(X, Y; \mu) - T(X, Y; \mu)) \stackrel{d}{\to} \int Q'_{X,Y} d\mu.$$

  where $Q'_{X,Y}$ is a centered Gaussian process.

- **Test of independence.**

Limit theory of distance covariance for triangular arrays

- Distance covariance between $(R_i, \mathbf{\Theta}_i)$ given $R_i > r_n$

$$\tilde{T}_n := T_n(R, \mathbf{\Theta}; \mu)\big|_{R > r_n}$$

- Effective sample size

$$k_n := \#\{R_i > r_n\}$$

Theorem

$$k_n \tilde{T}_n \xrightarrow{d} \int |\tilde{Q}|^2 d\mu,$$

where $\tilde{Q}$ is a centered Gaussian process.

Limit theory of distance covariance for triangular arrays

- Distance covariance between $(R_i, \mathbf{\Theta}_i)$ given $R_i > r_n$

$$\tilde{T}_n := T_n(R, \mathbf{\Theta}; \mu)\big|_{R>r_n}$$

- Effective sample size

$$k_n := \#\{R_i > r_n\}$$

- Note that $(R_i, \mathbf{\Theta}_i)\big|_{R_i>r_n}$, $r_n \to \infty$, $n \to \infty$, is a triangular array.

Theorem

Under suitable conditions,

$$k_n \tilde{T}_n \xrightarrow{d} \int |\tilde{Q}|^2 d\mu,$$

where $\tilde{Q}$ is a centered Gaussian process.

Limit theory of distance covariance for triangular arrays

$$k_n \tilde{T}_n \xrightarrow{d} \int |\tilde{Q}|^2 d\mu,$$

Sketch of the suitable conditions:

1. Effective sample size $k_n \to \infty$
   - thresholds $r_n \to \infty$ not too fast
2. $(R, \mathbf{\Theta})|_{R>r_n}$ becomes independent fast enough
   - thresholds $r_n \to \infty$ not too slow
3. conditions on weight measure μ such that reddistance covariance exists
   - since R is heavy-tailed
4. conditions on mixing coefficients α_h such that central limit theorem can be applied

Limit theory of distance covariance for triangular arrays

Details of the suitable conditions:

1. $n\mathbb{P}(R > r_n) \to \infty$;
2. $n\mathbb{P}(R > r_n) \int |\varphi_{\frac{R}{r_n},\Theta|r_n} - \varphi_{\frac{R}{r_n}|r_n}\varphi_{\Theta|r_n}|^2 d\mu \to 0$;
3. $\int (1 \wedge |s|^\beta)(1 \wedge |t|^2)\mu(ds, dt) < \infty$ for some $1 < \beta < 2 \wedge \alpha$;
4. there exists $l_n \to \infty$ such that $l_n\mathbb{P}(R > r_n) \to 0$ and
   a) $\mathbb{P}(R > r_n)^{-\delta} \sum_{h=l_n}^{\infty} \alpha_h^\delta \to 0$ for some $\delta \in (0, 1)$;
   b) $\lim_{h\to\infty} \limsup_{n\to\infty} \frac{1}{p_n} \sum_{j=h}^{l_n} \mathbb{P}(\|\mathbf{X}_0\| > r_n, \|\mathbf{X}_j\| > r_n) = 0$;
   c) $np_n\alpha_{l_n} \to 0$.

Limit theory of distance covariance for triangular arrays

Details of the suitable conditions:

1. $n\mathbb{P}(R > r_n) \to \infty$;
2. $n\mathbb{P}(R > r_n) \int |\varphi_{\frac{R}{r_n},\Theta|r_n} - \varphi_{\frac{R}{r_n}|r_n}\varphi_{\Theta|r_n}|^2 d\mu \to 0$;
3. $\int (1 \wedge |s|^\beta)(1 \wedge |t|^2)\mu(ds, dt) < \infty$ for some $1 < \beta < 2 \wedge \alpha$;
4. there exists $l_n \to \infty$ such that $l_n\mathbb{P}(R > r_n) \to 0$ and
   a) $\mathbb{P}(R > r_n)^{-\delta} \sum_{h=l_n}^{\infty} \alpha_h^\delta \to 0$ for some $\delta \in (0, 1)$;
   b) $\lim_{h\to\infty} \limsup_{n\to\infty} \frac{1}{p_n} \sum_{j=h}^{l_n} \mathbb{P}(\|\mathbf{X}_0\| > r_n, \|\mathbf{X}_j\| > r_n) = 0$;
   c) $np_n\alpha_{l_n} \to 0$.

Limit theory of distance covariance for triangular arrays

Details of the suitable conditions:

1. $n\mathbb{P}(R > r_n) \to \infty$;
2. $n\mathbb{P}(R > r_n) \int |\varphi_{\frac{R}{r_n},\Theta|r_n} - \varphi_{\frac{R}{r_n}|r_n}\varphi_{\Theta|r_n}|^2 d\mu \to 0$;
   - can be translated to a second-order RV type condition
3. $\int (1 \wedge |s|^\beta)(1 \wedge |t|^2)\mu(ds, dt) < \infty$ for some $1 < \beta < 2 \wedge \alpha$;
4. there exists $l_n \to \infty$ such that $l_n\mathbb{P}(R > r_n) \to 0$ and
   a) $\mathbb{P}(R > r_n)^{-\delta} \sum_{h=l_n}^{\infty} \alpha_h^\delta \to 0$ for some $\delta \in (0, 1)$;
   b) $\lim_{h\to\infty} \limsup_{n\to\infty} \frac{1}{p_n} \sum_{j=h}^{l_n} \mathbb{P}(\|\mathbf{X}_0\| > r_n, \|\mathbf{X}_j\| > r_n) = 0$;
   c) $np_n\alpha_{l_n} \to 0$.

Limit theory of distance covariance for triangular arrays

Details of the suitable conditions:

1. $n\mathbb{P}(R > r_n) \to \infty$;
2. $n\mathbb{P}(R > r_n) \int |\varphi_{\frac{R}{r_n},\Theta|r_n} - \varphi_{\frac{R}{r_n}|r_n}\varphi_{\Theta|r_n}|^2 d\mu \to 0$;
   - can be translated to a second-order RV type condition
3. $\int (1 \wedge |s|^{\beta})(1 \wedge |t|^2)\mu(ds, dt) < \infty$ for some $1 < \beta < 2 \wedge \alpha$;
   - can be relaxed if looking at $\log R$
4. there exists $l_n \to \infty$ such that $l_n\mathbb{P}(R > r_n) \to 0$ and
   a) $\mathbb{P}(R > r_n)^{-\delta} \sum_{h=l_n}^{\infty} \alpha_h^{\delta} \to 0$ for some $\delta \in (0, 1)$;
   b) $\lim_{h\to\infty} \limsup_{n\to\infty} \frac{1}{p_n} \sum_{j=h}^{l_n} \mathbb{P}(\|\mathbf{X}_0\| > r_n, \|\mathbf{X}_j\| > r_n) = 0$;
   c) $np_n\alpha_{l_n} \to 0$.

Limit theory of distance covariance for triangular arrays

Details of the suitable conditions:

1. $n\mathbb{P}(R > r_n) \to \infty$;
2. $n\mathbb{P}(R > r_n) \int |\varphi_{\frac{R}{r_n},\Theta|r_n} - \varphi_{\frac{R}{r_n}|r_n}\varphi_{\Theta|r_n}|^2 d\mu \to 0$;
   - can be translated to a second-order RV type condition
3. $\int (1 \wedge |s|^\beta)(1 \wedge |t|^2)\mu(ds, dt) < \infty$ for some $1 < \beta < 2 \wedge \alpha$;
   - can be relaxed if looking at $\log R$
4. there exists $l_n \to \infty$ such that $l_n\mathbb{P}(R > r_n) \to 0$ and
   a) $\mathbb{P}(R > r_n)^{-\delta} \sum_{h=l_n}^{\infty} \alpha_h^\delta \to 0$ for some $\delta \in (0, 1)$;
   b) $\lim_{h\to\infty} \limsup_{n\to\infty} \frac{1}{p_n} \sum_{j=h}^{l_n} \mathbb{P}(\|\mathbf{X}_0\| > r_n, \|\mathbf{X}_j\| > r_n) = 0$;
   c) $np_n\alpha_{l_n} \to 0$.
   - adapted from Davis & Mikosch (2009)

Illustration: $R \perp \Theta$ only when $R > r_{0.1}$, the upper 10%-quantile.

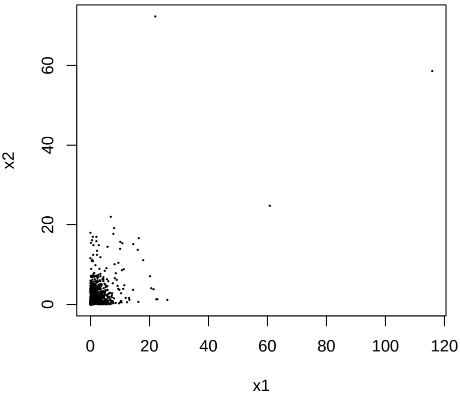


Illustration: $R \perp \Theta$ only when $R > r_{0.1}$

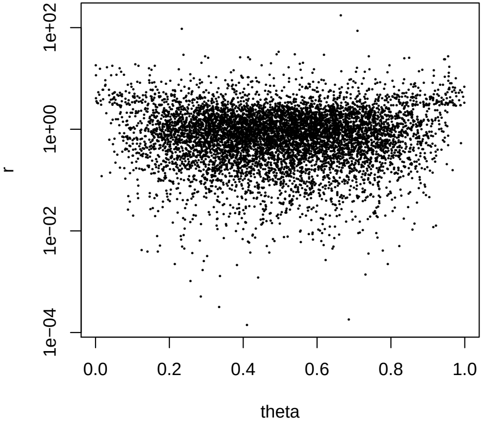


Illustration: $R \perp \Theta$ only when $R > r_{0.1}$

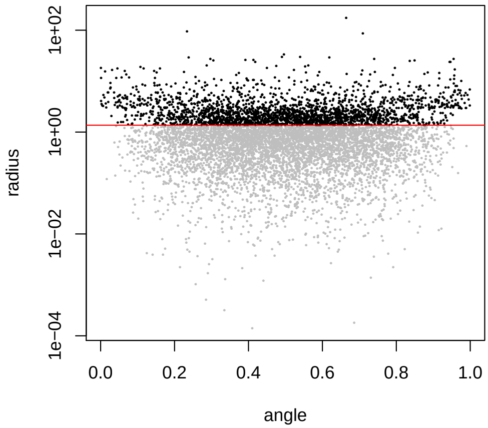


Illustration: $R \perp \Theta$ only when $R > r_{0.1}$

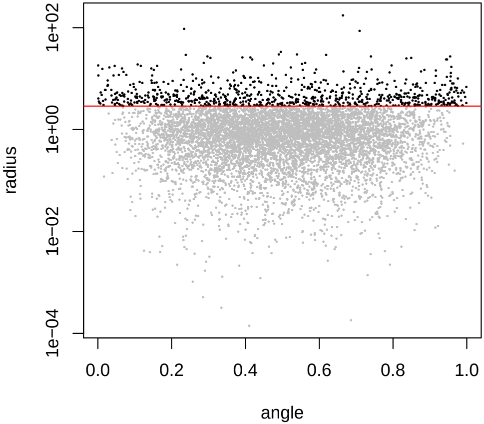


Illustration: $R \perp \Theta$ only when $R > r_{0.1}$

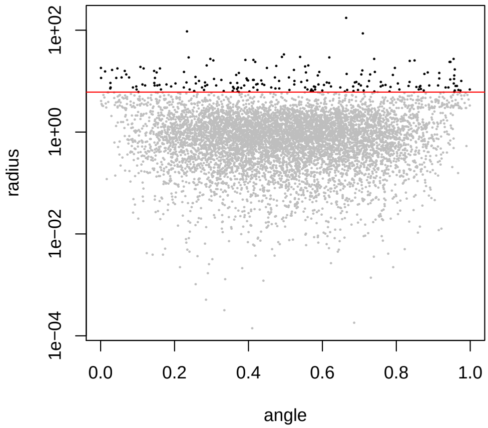


Illustration: $R \perp \Theta$ only when $R > r_{0.1}$

For each upper quantile r_q,

- calculate conditional distance covariance from $(R_{i_1}, \mathbf{\Theta}_{i_1}), \ldots, (R_{i_K}, \mathbf{\Theta}_{i_K})$ for which $R_{i_k} > r_q$
- derive the *p*-value of test of independence

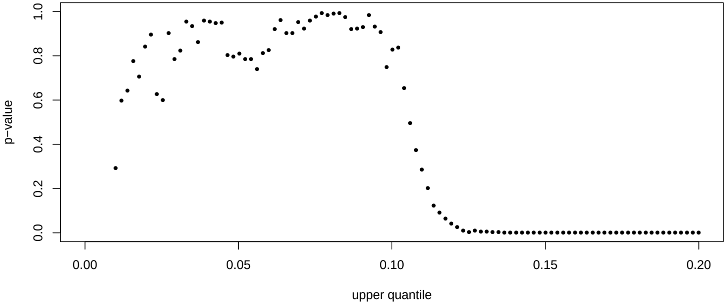


Illustration: $R \perp \Theta$ only when $R > r_{0.1}$

For each upper quantile r_q,

- calculate conditional distance covariance for m independent subsamples from $(R_{i_1}, \mathbf{\Theta}_{i_1}), \ldots, (R_{i_K}, \mathbf{\Theta}_{i_K})$ for which $R_{i_k} > r_q$
- derive the p-value of test of independence for each subsample

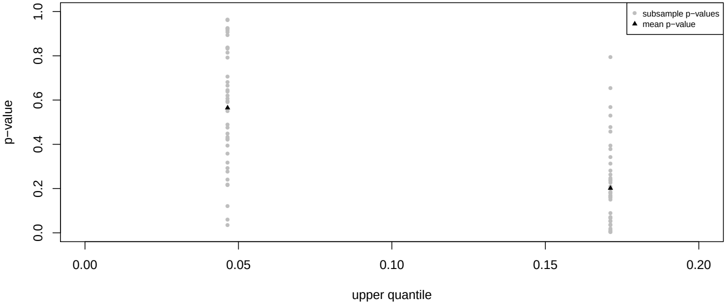


Illustration: $R \perp \Theta$ only when $R > r_{0.1}$

For each upper quantile r_q,

- calculate conditional distance covariance for m independent subsamples from $(R_{i_1}, \mathbf{\Theta}_{i_1}), \ldots, (R_{i_K}, \mathbf{\Theta}_{i_K})$ for which $R_{i_k} > r_q$
- derive the p-value of test of independence for each subsample
- average the p-values

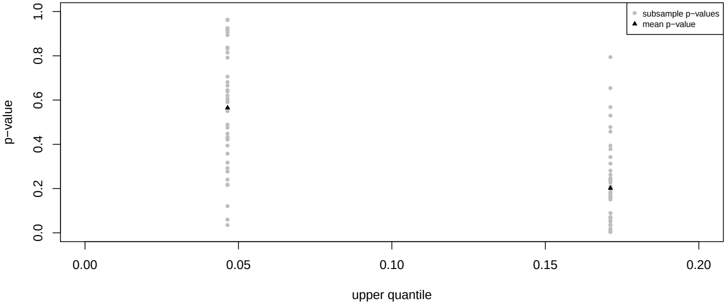


Illustration: $R \perp \Theta$ only when $R > r_{0.1}$

For each upper quantile r_q,

- calculate conditional distance covariance for m independent subsamples from $(R_{i_1}, \mathbf{\Theta}_{i_1}), \ldots, (R_{i_K}, \mathbf{\Theta}_{i_K})$ for which $R_{i_k} > r_q$
- derive the p-value of test of independence for each subsample
- average the p-values

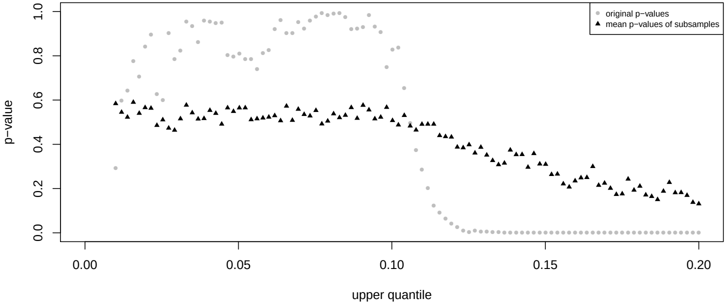
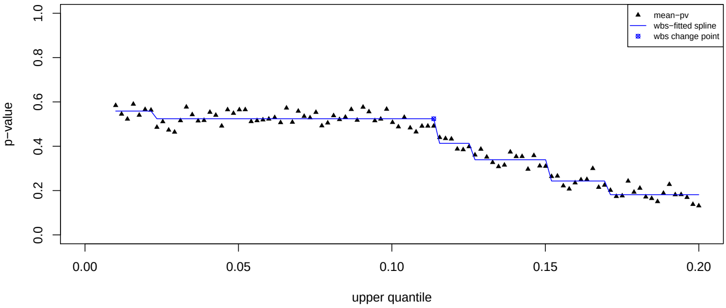


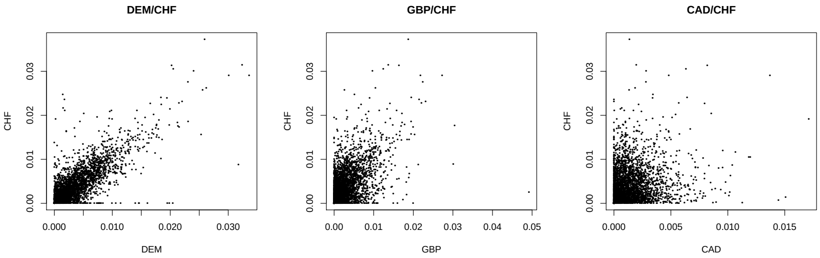
Illustration: $R \perp \Theta$ only when $R > r_{0.1}$

To choose the threshold,

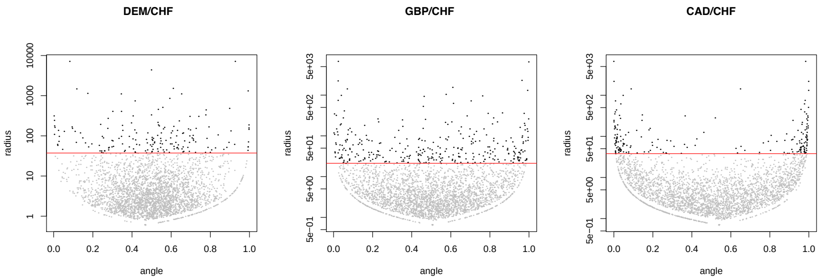
- when the mean of p-values falls below 0.5
- Wild Binary Segmentation (Fryzlewicz, 2014) fits a piecewise constant spline to the data based on CUSUM statistics



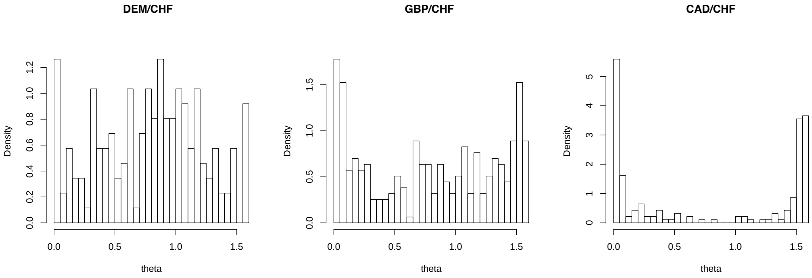
Daily absolute log-returns of exchange rates, from 1990-01-01 to 1998-12-31



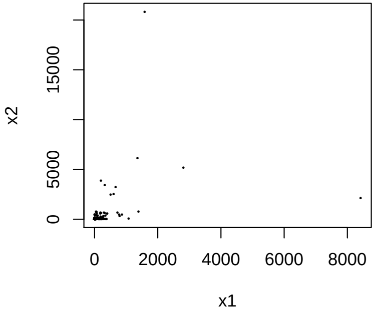
Daily absolute log-returns of exchange rates, from 1990-01-01 to 1998-12-31



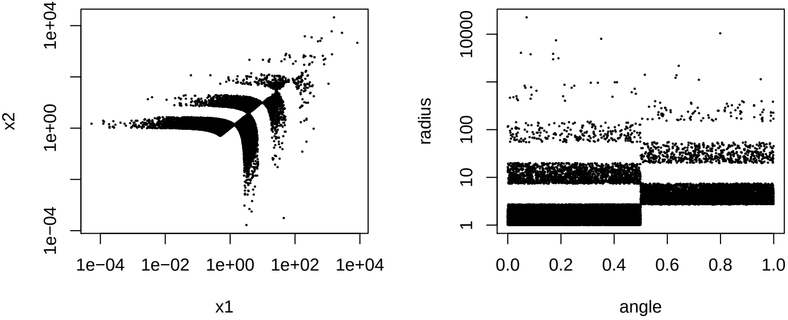
Daily absolute log-returns of exchange rates, from 1990-01-01 to 1998-12-31



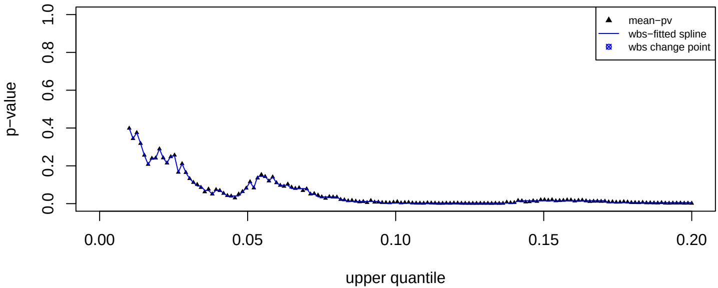
Detecting non-regular variation



Detecting non-regular variation



Detecting non-regular variation



Selected references

- Davis, R.A., Matsui, M., Mikosch T., Wan, P. (2018) Applications of Distance Correlation to Time Series. *Bernoulli.* **24**(4A), 3087–3116.
- Davis, R.A., Mikosch T. (2009) The extremogram: A correlogram for extreme events. *Bernoulli.* **15**(4), 977–1009.
- Feuerverger, A. (1993) A consistent test for bivariate dependence. *Int. Stat. Rev.* **61**, 419–433.
- Fryzlewicz, P. (2004) Wild binary segmentation for multiple change-point detection. *Ann. Statist.*, **42**(6):2243–2281.
- Meintanis, S.G. and Iliopoulos, G. (2008) Fourier methods for testing multivariate independence. *Comput. Statist. Data Anal.* **52**, 1884–1895.
- Székely, G.J., Rizzo, M.L. and Bakirov, N.K. (2007) Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35**, 2769–2794.
- Wan, P. and Davis, R.A. (2018+) Threshold selection for multivariate heavy-tailed data. *Extremes*.