

# A Appendix

- A.1 Risk in perspective
- A.2 Basics concepts in risk management
- A.4 Financial time series
- A.5 Extreme value theory
- A.6 Multivariate models
- A.7 Copulas and dependence
- A.8 Aggregate risk

# A.1 Risk in perspective

## Background information

- A *bond* is an instrument of indebtedness. The issuer owes the bond holder a debt and is obliged to pay at *maturity  $T$*  the principal and a *coupon* (interest; typically paid at fixed time points).
- *Netting* refers to the compensation of long versus short positions on the same underlying.
- A *derivative* is a financial instrument *derived from an underlying asset*, e.g. stocks, bonds, commodities, currencies, interest rates etc. Examples:
  - ▶ *Options* (*right, but not the obligation*, to buy (*call*) or sell (*put*) an asset at an agreed-upon price (the *strike price  $K$* ) during a predetermined period (*American*) or date (*exercise date  $T$* ; *European*);
  - ▶ *Futures* (*obligation* for the buyer (seller) to purchase (sell) an asset at a predetermined date and price);

- ▶ *Swaps* (any exchange of an asset for another to change the maturity (e.g. of a bond) or because investment objectives have changed; include currency swaps, interest rate swaps).
- A *credit default swap (CDS)* is a credit derivative which allows the (protection) buyer (who pays premiums) to transfer credit risk inherent in a reference entity to a seller (investor; pays in case of default).
- A *CDS spread* is the annual amount the protection buyer must pay the protection seller over  $[0, T]$ , expressed as a fraction (often in 1 *basis point* = 0.01%) of the notional amount.
- The *Fundamental Theorems of Asset Pricing*:
  - ▶ A (model for) a market is *arbitrage free* if and only if there exists a risk-neutral probability measure  $Q$  equivalent to  $\mathbb{P}$ ;
  - ▶ A market is *complete* (i.e. every contingent claim can be replicated) if and only if  $Q$  is unique.

## QRM beyond finance

- Some of the earliest applications of QRM are to be found in the **manufacturing industry**, where similar concepts and tools exist under names like **reliability** or **total quality control**. Industrial companies have recognized the **risks associated with bringing faulty products to the market**.
- QRM techniques have been adopted in the **transport and energy industries** (cost of storage and transport of electricity).
- There is an interest in the **transfer of risks between industries**; this process is known as **alternative risk transfer (ART)**, e.g. the risk transfer between the **insurance and banking industries**.
- QRM methodology also applies to **individuals**, e.g. via the **risk of unemployment**, **depreciation in the housing market** or the investment in the **education of children**.

## A.2 Basics concepts in risk management

### Background information

- A *balance sheet* is a financial statement showing *assets* (investments) and *liabilities* (obligations; show how funds have been raised)
- $(X_t)$  is a (discrete) *martingale* with respect to the filtration  $(\mathcal{F}_t)$  if
  - ▶  $X_t \in \mathcal{F}_t$  for all  $t \in \mathbb{N}_0$  (*adapted*);
  - ▶  $\mathbb{E}X_t < \infty$  for all  $t \in \mathbb{N}_0$ ;
  - ▶  $\mathbb{E}(X_{t+1} | \mathcal{F}_t) = X_t$  for all  $t \in \mathbb{N}_0$ .

## Physical ( $\mathbb{P}$ ) vs risk-neutral ( $\mathbb{Q}$ ) measure: An example

- Consider a defaultable bond with principal 1 and maturity  $T = 1y$ . In case of a default (real world probability  $p = 0.01$ ), the recovery rate is  $R = 60\%$ . The risk-free interest rate is  $r = 0.05$ . Moreover, assume the bond's current price to be  $V_0 = 0.941$  ( $t = 0$ ).
- The **expected discounted value** of the bond is

$$\frac{1}{1+r}(1 \cdot (1-p) + R \cdot p) = \frac{1}{1.05}(0.99 + 0.6p) = 0.949$$

which is  $> V_0$  since investors demand a **premium** for bearing the bond's **default risk**.

- Here,  $\mathbb{Q}$  is determined by specifying a  $q$  such that

$$\frac{1}{1+r}(1 \cdot (1-q) + R \cdot q) = V_0.$$

This implies  $q = 0.03$  which is greater than  $p = 0.01$ ; the larger value reflects the risk premium.

## Other formulas for expected shortfall

### Proposition A.1 (ES formulas)

For  $\alpha \in (0, 1)$ ,

$$\begin{aligned} 1) \quad \text{ES}_\alpha(L) &= \frac{\mathbb{E}((L - F_L^{\leftarrow}(\alpha))_+)}{1 - \alpha} + F_L^{\leftarrow}(\alpha); \\ 2) \quad \text{ES}_\alpha(L) &= \frac{\mathbb{E}(LI_{\{L > F_L^{\leftarrow}(\alpha)\}}) + F_L^{\leftarrow}(\alpha)(1 - \alpha - \bar{F}_L(F_L^{\leftarrow}(\alpha)))}{1 - \alpha}. \end{aligned}$$

*Proof of Proposition A.1.*

1) Recall that  $L \stackrel{d}{=} F_L^{\leftarrow}(U)$ ,  $U \sim \text{U}(0, 1)$ . Therefore,

$$\begin{aligned} \frac{\mathbb{E}((L - F_L^{\leftarrow}(\alpha))_+)}{1 - \alpha} &= \frac{1}{1 - \alpha} \int_0^1 (F_L^{\leftarrow}(u) - F_L^{\leftarrow}(\alpha))_+ du \\ &= \frac{1}{1 - \alpha} \int_\alpha^1 (F_L^{\leftarrow}(u) - F_L^{\leftarrow}(\alpha)) du \\ &= \text{ES}_\alpha(L) - F_L^{\leftarrow}(\alpha). \end{aligned}$$

2) First note that

$$\begin{aligned}
 \mathbb{E}((L - F_L^{\leftarrow}(\alpha))_+) &= \mathbb{E}((L - F_L^{\leftarrow}(\alpha))I_{\{L > F_L^{\leftarrow}(\alpha)\}}) \\
 &= \mathbb{E}(LI_{\{L > F_L^{\leftarrow}(\alpha)\}}) - F_L^{\leftarrow}(\alpha)\mathbb{E}(I_{\{L > F_L^{\leftarrow}(\alpha)\}}) \\
 &= \mathbb{E}(LI_{\{L > F_L^{\leftarrow}(\alpha)\}}) - F_L^{\leftarrow}(\alpha)\bar{F}_L(F_L^{\leftarrow}(\alpha)).
 \end{aligned}$$

Now apply 1), divide by  $1 - \alpha$  and add  $F_L^{\leftarrow}(\alpha)$ . □

### Corollary A.2 (ES formulas under continuous $F_L$ )

Let  $F_L$  be continuous at  $F_L^{\leftarrow}(\alpha)$ . Then

- 1)  $\text{ES}_\alpha(L) = \frac{\mathbb{E}(LI_{\{L > F_L^{\leftarrow}(\alpha)\}})}{1 - \alpha}$
- 2)  $\text{ES}_\alpha(L) = \mathbb{E}(L \mid L > F_L^{\leftarrow}(\alpha))$  (i.e. *conditional VaR (CVaR)*)

*Proof of Corollary A.2.*

- 1) Since  $\bar{F}_L(F_L^{\leftarrow}(\alpha)) = 1 - F_L(F_L^{\leftarrow}(\alpha)) = 1 - \alpha$  for all  $\alpha \in \text{ran } F_L \cup \{\inf F_L, \sup F_L\} \supseteq (0, 1)$ , the claim follows from Proposition A.1 2).



2) First note that

$$\begin{aligned} F_{L|L>F_L^{\leftarrow}(\alpha)}(x) &= \mathbb{P}(L \leq x \mid L > F_L^{\leftarrow}(\alpha)) = \frac{\mathbb{P}(F_L^{\leftarrow}(\alpha) < L \leq x)}{\mathbb{P}(L > F_L^{\leftarrow}(\alpha))} \\ &= \frac{F_L(x) - F_L(F_L^{\leftarrow}(\alpha))}{1 - F_L(F_L^{\leftarrow}(\alpha))} I_{\{x > F_L^{\leftarrow}(\alpha)\}} = \frac{F_L(x) - \alpha}{1 - \alpha} I_{\{x > F_L^{\leftarrow}(\alpha)\}}, \end{aligned}$$

where the latter equality holds since  $\alpha \in \text{ran } F_L$ . This implies

$$\begin{aligned} \mathbb{E}(L \mid L > F_L^{\leftarrow}(\alpha)) &= \int_{\mathbb{R}} x dF_{L|L>F_L^{\leftarrow}(\alpha)}(x) = \int_{F_L^{\leftarrow}(\alpha)}^{\infty} x \frac{dF_L(x)}{1 - \alpha} \\ &= \frac{\mathbb{E}(L I_{\{L > F_L^{\leftarrow}(\alpha)\}})}{1 - \alpha} \stackrel{1)}{=} \text{ES}_{\alpha}(L). \end{aligned} \quad \square$$

## Elicitability explained in words

We follow Kou and Peng (2014, Sections 1 and 2.2) and McNeil et al. (2005, Chapter 9).

- Computing a (one-period ahead) risk measure  $\varrho(L) =: \varrho(F_L)$  is a point forecasting problem because  $F_L$  is unknown and one has to find an estimate  $\hat{F}_L$  of it and forecast the unknown true  $\varrho(F_L)$  via the point forecast  $\varrho(\hat{F}_L)$ .
- As different  $\hat{F}_L$  can be used to forecast the risk measure, it is desirable to be able to evaluate which of them gives a better point forecast.
- Suppose we want to forecast  $L$  (or  $F_L$ ) by a point  $y$ . The *forecasting error* is

$$\mathbb{E}(S(y, L)) = \int_{\mathbb{R}} S(y, l) dF_L(l),$$

where  $S(y, l)$  is a *scoring* (i.e. forecasting objective) *function*.

- Two point forecasting methods can be compared via their forecasting errors. For a given  $S$ , the **optimal point forecast** is

$$\varrho^*(F_L) = \operatorname{arginf}_y \mathbb{E}(S(y, L)) \quad (\text{minimizing the forecast error}).$$

For example, for  $S(y, l) = (y - l)^2$  and  $S(y, l) = |y - l|$ , the optimal point forecasts are the mean and median of  $F_L$ , respectively.

- Elicitable risk measures** (or: statistical functionals) **are risk measures  $\varrho$  which minimize  $\mathbb{E}(S(y, L))$  of some scoring function  $S$** ; hence that  $S$  can be used to compare different point forecasting procedures for  $\varrho$  (“the smaller the forecasting error, the better” makes sense).
- If  $\varrho$  is not elicitable, one cannot find such an  $S$  and thus the minimization of the forecasting error does not yield the true value  $\varrho(F_L)$  for any  $S$ .** Hence, for two competing point forecasts of  $\varrho(F_L)$ , one cannot tell which performs the best by comparing their forecasting error, no matter what  $S$  is used.

## The (nonparametric) bootstrap

- Suppose  $X_1, \dots, X_n \stackrel{\text{ind.}}{\sim} F$  ( $F$  unknown) and we are interested in estimating  $\theta = \theta(F)$ . If we could at least sample from  $F$ , we would estimate  $\theta$  by  $\hat{\theta}_n = \theta(\hat{F}_n)$ . But even then, we would only get a point estimate  $\hat{\theta}_n$  of  $\theta$ . How do we get a  $(1 - \beta)$ -confidence interval for  $\theta$ ? How  $\text{var}(\hat{\theta}_n)$ ? How the distribution of  $\hat{\theta}_n$ ?
- The (nonparametric) bootstrap treats  $\hat{F}_n$  as the true df  $F$  (first approximation) and samples from that (i.e. resamples  $X_1, \dots, X_n$ ) to approximate the distribution of  $\hat{\theta}_n$  (second approximation).
- Advantages: Applicable if  $F$  or the df of  $\theta$  is unknown; applicable if  $n$  is small (unlike the CLT); applicable if  $F$  is skewed (CIs based on the CLT are always centered around the sample mean); easy to implement.

### Algorithm A.3 (Nonparametric bootstrap)

- 1) Fix a large  $B \in \mathbb{N}$ .

- 2) For  $b \in \{1, \dots, B\}$ , do:
  - 2.1) Randomly sample  $X_{b,1}, \dots, X_{b,n}$  from  $X_1, \dots, X_n$  with replacement.
  - 2.2) Compute the bootstrap estimator  $\hat{\theta}_{b,n} = \theta(\hat{F}_{b,n})$  where  $\hat{F}_{b,n}(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_{b,i} \leq x\}}$  (informally: compute  $\hat{\theta}_{b,n}$  from  $X_{b,1}, \dots, X_{b,n}$ ).
- 3) Use the *bootstrap sample*  $\hat{\theta}_{b,n}$ ,  $b \in \{1, \dots, B\}$ , to approximate the distribution of  $\hat{\theta}_n$ .

### Examples for Step 3):

- $\hat{\mu}_{B,n} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_{b,n} \approx \mathbb{E}(\hat{\theta}_n)$ ;  $\hat{\sigma}_{B,n}^2 = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_{b,n} - \hat{\mu}_{B,n})^2 \approx \text{var}(\hat{\theta}_n)$ ; the empirical df based on the  $\hat{\theta}_{b,n}$ 's approximates the df of  $\hat{\theta}_n$ .
- An *bootstrapped*  $(1 - \beta)$ -*confidence interval* for  $\theta$  based on  $\hat{\theta}_{(1),n} \leq \dots \leq \hat{\theta}_{(B),n}$  is

$$\left[ \hat{\theta}_{(\max\{\lfloor \frac{\beta}{2} B \rfloor, 1\}),n}, \hat{\theta}_{(\lceil (1-\frac{\beta}{2}) B \rceil),n} \right].$$

## A.4 Financial time series

### Conditional expectations

#### Definition A.4 (Conditional expectation, conditional probability)

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space,  $\mathbf{X} \in L^1(\Omega, \mathcal{F}, \mathbb{P})$  – i.e.  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^d$ ,  $\mathbf{X}$  is  $\mathcal{F}$ -measurable (i.e.  $\mathbf{X}^{-1}(B) \in \mathcal{F}$  for all  $B \in \mathcal{B}(\mathbb{R}^d)$ ) and  $\mathbb{E}|\mathbf{X}| < \infty$  – and  $\mathcal{G} \subseteq \mathcal{F}$  be a  $\sigma$ -algebra. Then any rv  $\mathbf{Y}$  such that

1)  $\mathbf{Y} \in \mathcal{G}$  ( $\mathbf{Y}$  is  $\mathcal{G}$ -measurable);

2)  $\mathbb{E}|\mathbf{Y}| < \infty$ ; and

3)  $\mathbb{E}(\mathbf{Y} I_G) = \int_G \mathbf{Y} d\mathbb{P} = \int_G \mathbf{X} d\mathbb{P} = \mathbb{E}(\mathbf{X} I_G)$  for all  $G \in \mathcal{G}$

is called *conditional expectation of  $\mathbf{X}$  given  $\mathcal{G}$*  and denoted by  $\mathbb{E}(\mathbf{X} | \mathcal{G})$ .

$\mathbb{P}(A | \mathcal{G}) = \mathbb{E}(I_A | \mathcal{G})$  is called *conditional probability of  $A$  given  $\mathcal{G}$* .

The following property of conditional expectations is used frequently and known as *tower property*.

### Lemma A.5 (Tower property; the smallest $\sigma$ -algebra remains)

If  $\mathcal{G} \subseteq \mathcal{F}$ , then  $\mathbb{E}(\mathbb{E}(X | \mathcal{G}) | \mathcal{F}) = \mathbb{E}(X | \mathcal{G}) = \mathbb{E}(\mathbb{E}(X | \mathcal{F}) | \mathcal{G})$ .

*Idea of proof.* Let  $G \in \mathcal{G} \subseteq \mathcal{F}$ . Applying Definition A.4 Part 3) to  $\mathbb{E}(\mathbb{E}(X | \mathcal{G}) | \mathcal{F})$  and then to  $\mathbb{E}(X | \mathcal{G})$  implies that  $\mathbb{E}(\mathbb{E}[\mathbb{E}[X | \mathcal{G}] | \mathcal{F}]I_G) = \mathbb{E}(\mathbb{E}[X | \mathcal{G}]I_G) = \mathbb{E}(XI_G)$ .  $\square$

## On partial autocorrelation in stationary time series

For introducing it, we need some tools.

- *Hilbert's Projection Theorem* (see Brockwell and Davis (1991, p. 51)):

If  $\mathcal{H}_0$  is a closed subspace of the Hilbert space  $\mathcal{H}$  and  $x \in \mathcal{H}$ , then:

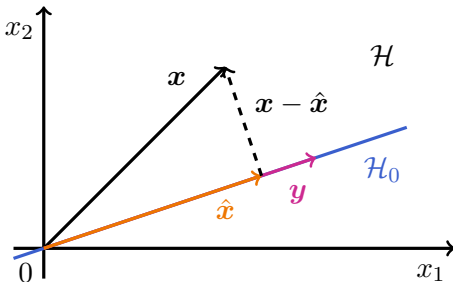
- There exists a unique  $\hat{x} \in \mathcal{H}_0 : \|x - \hat{x}\| = \inf_{y \in \mathcal{H}_0} \|x - y\|$ ;
- $\hat{x} \in \mathcal{H}_0$ ,  $\|x - \hat{x}\| = \inf_{y \in \mathcal{H}_0} \|x - y\|$  if and only if  $\hat{x} \in \mathcal{H}_0$ ,  $x - \hat{x} \in \mathcal{H}_0^\perp = \{x \in \mathcal{H} : \langle x, y \rangle = 0 \text{ for all } y \in \mathcal{H}_0\}$ .

### Note:

- ▶  $\hat{x}$  is the (orthogonal) projection of  $x$  onto  $\mathcal{H}_0$ , denoted by  $P_{\mathcal{H}_0}x$ .
- ▶  $\hat{x} = P_{\mathcal{H}_0}x$  is the unique element:  $\langle x - \hat{x}, y \rangle = 0 \ \forall y \in \mathcal{H}_0$  (prediction equations;  $P_{\mathcal{H}_0}x$  is the best approximation/prediction of  $x$  in  $\mathcal{H}_0$ ).

### Example A.6

$$x \in \mathcal{H} = \mathbb{R}^2, \mathcal{H}_0 = \text{span}\{y\}$$





- **Yule–Walker equations.** Let  $X_1, \dots, X_{n-1}, X_n$  be elements of a stationary time series  $(X_t)_{t \in \mathbb{Z}}$  with  $\mu(t) = 0$ ,  $t \in \mathbb{Z}$ . Suppose we would like to find  $\hat{X}_n = \sum_{k=1}^{n-1} \phi_{n-1,k} X_{n-k}$  such that

$$\mathbb{E}((X_n - \hat{X}_n)^2) \rightarrow \min_{(\phi_{n-1,k})_{k=1}^{n-1}}.$$

$\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathbb{P})$  is a Hilbert space with  $\langle X, Y \rangle = \mathbb{E}(XY)$  and  $\mathcal{H}_{n-1} = \text{span}\{X_1, \dots, X_{n-1}\} = \{\sum_{k=1}^{n-1} \alpha_k X_{n-k} : \alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}\}$  is a subspace. Therefore,  $\hat{X}_n = P_{\mathcal{H}_{n-1}} X_n$  satisfies the prediction equations

$$\begin{aligned} \langle X_n - \hat{X}_n, Y \rangle &= 0, \quad \forall Y \in \mathcal{H}_{n-1} \\ \Leftrightarrow \langle X_n - \hat{X}_n, \sum_{k=1}^{n-1} \alpha_k X_{n-k} \rangle &= 0, \quad \forall \alpha_1, \dots, \alpha_{n-1} \in \mathbb{R} \\ &= \underbrace{\sum_{k=1}^{n-1} \alpha_k \langle X_n - \hat{X}_n, X_{n-k} \rangle} \end{aligned}$$

$$\begin{aligned}
&\Leftrightarrow \underbrace{\langle X_n - \hat{X}_n, X_l \rangle}_{=0} = 0, \quad \forall l \in \{1, \dots, n-1\} \\
&= \mathbb{E}((X_n - \sum_{k=1}^{n-1} \phi_{n-1,k} X_{n-k}) X_l) \\
&= \mathbb{E}(X_n X_l) - \sum_{k=1}^{n-1} \phi_{n-1,k} \mathbb{E}(X_{n-k} X_l) \\
&\Leftrightarrow \gamma(n-l) = \sum_{k=1}^{n-1} \gamma(n-k-l) \phi_{n-1,k} \\
&\stackrel{\text{station.}}{\Leftrightarrow} \gamma(h) = \sum_{k=1}^{n-1} \gamma(h-k) \phi_{n-1,k}, \quad \forall h \in \{1, \dots, n-1\} \\
&\Leftrightarrow \Gamma_{n-1} \phi_{n-1} = \gamma_{n-1}, \quad (\text{Yule-Walker equations})
\end{aligned}$$

where

$$\begin{aligned}
\phi_{n-1} &= (\phi_{n-1,1}, \dots, \phi_{n-1,n-1}), \\
\gamma_{n-1} &= (\gamma(1), \dots, \gamma(n-1)), \\
\Gamma_{n-1} &= (\gamma(|i-j|))_{i,j=1}^{n-1}.
\end{aligned}$$

Hilbert's Projection Theorem ii)  $\Rightarrow$  there exists at least one solution  $\phi_{n-1}$  and all of them lead to the same  $\hat{X}_n$  (unique by i)). If  $\Gamma_{n-1}$

is regular (invertible),  $\phi_{n-1}$  is unique. This holds, e.g. if  $\gamma(0) > 0$ ,  $\gamma(h) \rightarrow 0$  ( $h \rightarrow \infty$ ); see Brockwell and Davis (1991, p. 167).

- $\phi_n$  can be computed with the *Durbin–Levinson algorithm*: Let  $(X_t)_{t \in \mathbb{Z}}$  be stationary with  $\mu(t) = 0$ ,  $t \in \mathbb{Z}$ ,  $\gamma(0) > 0$ ,  $\gamma(h) \rightarrow 0$  ( $h \rightarrow \infty$ ). Then, for all  $n \in \mathbb{N}$ ,

$$\begin{aligned}\phi_{n,n} &= \frac{\gamma(n) - \sum_{k=1}^{n-1} \gamma(n-k) \phi_{n-1,k}}{(*) \gamma(0) - \sum_{k=1}^{n-1} \gamma(n-k) \phi_{n-1,n-k}} \\ &= \frac{\rho(n) - \sum_{k=1}^{n-1} \rho(n-k) \phi_{n-1,k}}{1 - \sum_{k=1}^{n-1} \rho(n-k) \phi_{n-1,n-k}},\end{aligned}$$

$$\begin{pmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} \stackrel{(**)}{=} \begin{pmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix} - \phi_{n,n} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix}.$$

*Proof.* The Yule–Walker equations hold if and only if

$$\begin{pmatrix} \gamma(0) & \cdots & \gamma(n-2) & \gamma(n-1) \\ \vdots & \ddots & \vdots & \vdots \\ \cdots & \cdots & \gamma(0) & \vdots \\ \cdots & \cdots & \cdots & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n-1} \\ \phi_{n,n} \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n-1) \\ \gamma(n) \end{pmatrix}$$

$$\Leftrightarrow \Gamma_{n-1} \begin{pmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} + \underbrace{\phi_{n,n} \begin{pmatrix} \gamma(n-1) \\ \vdots \\ \gamma(1) \end{pmatrix}}_{\stackrel{=}{\Gamma_{n-1}} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix}} = \underbrace{\begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n-1) \end{pmatrix}}_{\stackrel{=}{\Gamma_{n-1}} \begin{pmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix}},$$

and  $\sum_{k=1}^{n-1} \gamma(n-k)\phi_{n,k} + \phi_{n,n}\gamma(0) \stackrel{(***)}{=} \gamma(n)$ . Multiplying with  $\Gamma_{n-1}^{-1}$  leads (\*\*). For (\*), use the  $k$ th row  $\phi_{n,k} = \phi_{n-1,k} - \phi_{n,n}\phi_{n-1,n-k}$  in (\*\*\*) and solve w.r.t.  $\phi_{n,n}$ .  $\square$

## Definition A.7 (PACF)

The *partial autocorrelation function (PACF)* of a stationary time series  $(X_t)_{t \in \mathbb{Z}}$  with  $\mu(t) = 0$ ,  $t \in \mathbb{Z}$ ,  $\gamma(0) > 0$ ,  $\gamma(h) \rightarrow 0$  ( $h \rightarrow \infty$ ) is

$$\begin{aligned}\phi(h) &= \text{corr}(X_0 - P_{\mathcal{H}_{h-1}}X_0, X_h - P_{\mathcal{H}_{h-1}}X_h) \\ &= \frac{\mathbb{E}(X_0(X_h - P_{\mathcal{H}_{h-1}}X_h))}{\mathbb{E}((X_h - P_{\mathcal{H}_{h-1}}X_h)(X_h - P_{\mathcal{H}_{h-1}}X_h))} \\ &= \frac{\mathbb{E}(X_0X_h) - \sum_{k=1}^{h-1} \phi_{h-1,k} \mathbb{E}(X_0X_{h-k})}{\mathbb{E}(X_h(X_h - P_{\mathcal{H}_{h-1}}X_h))} \\ &= \frac{\gamma(h) - \sum_{k=1}^{h-1} \gamma(h-k) \phi_{h-1,k}}{\text{station. } \gamma(0) - \sum_{k=1}^{h-1} \gamma(k) \phi_{h-1,k}} \stackrel{\text{DL algorithm}}{=} \phi_{h,h}, \quad h \in \mathbb{Z}.\end{aligned}$$

- PACF for MA(1): Let  $\theta = \theta_1$ .

$$\rho(h) = \begin{cases} 1, & \text{if } h = 0, \\ \frac{\theta}{1+\theta^2}, & \text{if } |h| = 1, \\ 0, & \text{if } |h| > 1. \end{cases}$$

Yule–Walker equations  $\Leftrightarrow P_h \phi_h = \rho_h$ . One can show by induction (or the Durbin–Levinson algorithm) that

$$\phi_{h,h} = -\frac{(-\theta)^h(1-\theta^2)}{1-\theta^{2(h+1)}}, \quad h \in \mathbb{N},$$

$$\phi_{h,h-k} = (-\theta)^{-k} \left( \frac{1-\theta^{2(k+1)}}{1-\theta^2} \right) \phi_{h,h}, \quad k \in \{1, \dots, h-1\}.$$

In particular,  $\phi(h) = \phi_{h,h} \searrow 0$  exponentially.

- PACF for AR( $p$ ): For  $h > p$ , let  $Y \in \mathcal{H}_{h-1} = \text{span}\{X_1, \dots, X_{h-1}\}$ . Since  $(X_t)_{t \in \mathbb{Z}}$  is causal,  $Y \in \text{span}\{\varepsilon_s : s \leq h-1\}$ . Thus,

$$\left\langle X_h - \sum_{k=1}^p \phi_k X_{h-k}, Y \right\rangle = \langle \varepsilon_t, Y \rangle = 0.$$

Prediction equations  $\Rightarrow \sum_{k=1}^p \phi_k X_{h-k}$  is the best linear approximation in the  $L^2$ -sense to  $X_h$  from  $X_1, \dots, X_{h-1}$ , so  $\sum_{k=1}^p \phi_k X_{h-k} = P_{\mathcal{H}_{h-1}} X_h$ . Hence,

$$\phi(h) = \text{corr}(\underbrace{X_0 - P_{\mathcal{H}_{h-1}} X_0}_{\in \text{span}\{X_0, \dots, X_{h-1}\}}, \underbrace{X_h - P_{\mathcal{H}_{h-1}} X_h}_{=\varepsilon_h}) \stackrel{\text{causality}}{=} 0.$$

### *Proof idea of Theorem 4.10.*

“ $\Leftarrow$ ”  $\phi(z) \neq 0$ ,  $|z| \leq 1 \Rightarrow 1/\phi(z)$  holomorphic on  $|z| < 1 + \varepsilon$  for some  $\varepsilon > 0 \Rightarrow 1/\phi(z) = \sum_{k=0}^{\infty} a_k z^k$ ,  $a_k(1 + \varepsilon/2)^k \rightarrow 0$  ( $k \rightarrow \infty$ )  
 $\Rightarrow \exists c > 0 : |a_k| < c(1 + \varepsilon/2)^{-k}$ ,  $k \in \mathbb{N}_0 \Rightarrow \sum_{k=0}^{\infty} |a_k| < \infty$ .  
 Proposition 4.9  $\Rightarrow \varepsilon_t/\phi(B)$  is stationary.  $\phi(B)X_t = \theta(B)\varepsilon_t \Rightarrow$   
 $X_t = \frac{1}{\phi(B)}\phi(B)X_t = \theta(B)\varepsilon_t/\phi(B)$  is stationary (and causal).

“ $\Rightarrow$ ”  $X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} = \psi(B)\varepsilon_t$ ,  $\sum_{k=0}^{\infty} |\psi_k| < \infty \Rightarrow \theta(B)\varepsilon_t =$   
 $\phi(B)X_t = \eta(B)\varepsilon_t$  for  $\eta(B) = \phi(B)\psi(B)$ . Let  $\eta(z) = \phi(z)\psi(z) =$   
 $\sum_{k=0}^{\infty} \eta_k z^k$ ,  $|z| \leq 1$ . With  $\theta_0 = 1$ , it follows that  $\sum_{k=0}^q \theta_k \varepsilon_{t-k} =$   
 $\sum_{k=0}^{\infty} \eta_k \varepsilon_{t-k}$ . Applying  $\mathbb{E}(\cdot \varepsilon_{t-j})$  ( $\langle \cdot, \varepsilon_{t-j} \rangle$ ) and using that  $(\varepsilon_t) \sim$   
 $\text{WN}(0, \sigma^2)$ , we obtain  $\eta_k = \theta_k$ ,  $k \in \{0, \dots, q\}$ , and  $\eta_k = 0$ ,  $k > q$ .  
 This implies that  $\theta(z) = \eta(z) = \phi(z)\psi(z)$  for all  $|z| \leq 1$ . Assume  
 $\phi(z_0) = 0$  for some  $|z_0| \leq 1$ . Then  $0 \neq \theta(z_0) = 0 \cdot \psi(z_0)$ . Since  
 $|\psi(z)| \leq \sum_{k=0}^{\infty} |\psi_k| < \infty$  for all  $|z| \leq 1$ , we obtain a contradiction.  
 Thus  $\phi(z) \neq 0$  for all  $|z| \leq 1$ . □



## Properties of (Q)MLEs

- We consider **two situations**: The model which has been fitted...
  - 1) ... has been **correctly specified**;
  - 2) ... has the correct dynamics but the **innovation distribution is erroneously assumed to be Gaussian** (in this case the MLE is known as **quasi-maximum likelihood estimator (QMLE)**).
- The **asymptotic results** for GARCH models are **similar to the results in the iid case**; they have been derived in a series of papers. We only treat pure GARCH models, the form of the results will apply more generally (e.g. to ARMA models with GARCH errors).
- Under 1), one can show that for a **GARCH( $p, q$ )** model with Gaussian innovations,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow[(n \rightarrow \infty)]{d} N_{p+q+1}(\mathbf{0}, I(\boldsymbol{\theta})^{-1}),$$

where

$$I(\boldsymbol{\theta}) := \mathbb{E}\left(\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left(\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)'\right) = -\mathbb{E}\left(\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\right) =: J(\boldsymbol{\theta})$$

is the *Fisher (or: expected) information* matrix. Thus we have a *consistent and asymptotically normal estimator*.

- In practice, the  $I(\boldsymbol{\theta})$  is often approximated by an *observed information matrix*. Two candidates are

$$\bar{I}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left(\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)'\right) \quad \text{and} \quad \bar{J}(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2},$$

where the former has *outer-product* and the latter has *Hessian* form. Evaluating them at the MLEs leads to  $\bar{I}(\hat{\boldsymbol{\theta}}_n)$  or  $\bar{J}(\hat{\boldsymbol{\theta}}_n)$ ; in practice, the derivatives are often approximated using first and second-order differences. Under 1),  $\bar{I}(\hat{\boldsymbol{\theta}}_n) \approx \bar{J}(\hat{\boldsymbol{\theta}}_n)$ . One could also take the *sandwich estimator*  $\bar{J}(\hat{\boldsymbol{\theta}}_n) \bar{I}(\hat{\boldsymbol{\theta}}_n)^{-1} \bar{J}(\hat{\boldsymbol{\theta}}_n)$ .

- Under 2), one *still obtains a consistent estimator*. If the true innovation

distribution has finite fourth moment, then

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow[(n \rightarrow \infty)]{d} N_{p+q+1}(\mathbf{0}, J(\boldsymbol{\theta})^{-1} I(\boldsymbol{\theta}) J(\boldsymbol{\theta})^{-1}),$$

Note that  $I(\boldsymbol{\theta})$  and  $J(\boldsymbol{\theta})$  typically differ in this case.  $J(\boldsymbol{\theta})^{-1} I(\boldsymbol{\theta}) J(\boldsymbol{\theta})^{-1}$  can be **estimated by the sandwich estimator**.

- If model checking suggests that the **dynamics** have been **adequately described** by the GARCH model, but **the Gaussian assumption seems doubtful**, then **standard errors for parameter estimates should be computed based on this covariance matrix estimate**.

## The SARIMA model

$(X_t)_{t \in \mathbb{Z}}$  is a SARIMA( $p, d, q$ )  $\times$  ( $\tilde{p}, \tilde{d}, \tilde{q}$ )<sub>s</sub> (Seasonal; Integrated) process if

$$\underbrace{\phi(B)}_{\text{order } p} \underbrace{\tilde{\phi}(B^s)}_{\text{order } s\tilde{p}} \overbrace{(1-B)^d (1-B^s)^{\tilde{d}}}^{\text{integrated part}} X_t = \underbrace{\theta(B)}_{\text{order } q} \underbrace{\tilde{\theta}(B^s)}_{\text{order } s\tilde{q}} \varepsilon_t, \quad t \in \mathbb{Z}.$$

We see that this is also an ARMA( $d + p + s(\tilde{d} + \tilde{p}), q + s\tilde{q}$ ) process. (Seasonal) “differences” are taken to get data from a **stationary** model.

## A.5 Extreme value theory

### The convergence to types theorem

#### Theorem A.8 (Convergence to types)

Suppose  $(M_n)_n$  is a sequence of rvs such that  $\frac{M_n - d_n}{c_n} \xrightarrow{d} Y$  for a rv  $Y$  and  $d_n \in \mathbb{R}$ ,  $c_n > 0$ . Then

$$\frac{M_n - \delta_n}{\gamma_n} \xrightarrow{d} Z$$

for a rv  $Z$  and  $\delta_n \in \mathbb{R}$ ,  $\gamma_n > 0$  if and only if

$$(c_n/\gamma_n) \rightarrow c \in [0, \infty), \quad (d_n - \delta_n)/\gamma_n \rightarrow d \in \mathbb{R},$$

in which case  $Z \stackrel{d}{=} cY + d$  (i.e.  $Y$  and  $Z$  are of the same type) and  $c, d$  are the unique such constants.

*Proof.* See Embrechts, Klüppelberg, et al. (1997, p. 554). □

# The Gumbel MDA

## Theorem A.9 (Gumbel MDA)

$F \in \text{MDA}(H_0)$  if and only if there exists  $z < x_F \leq \infty$  such that

$$\bar{F}(x) = c(x) \exp\left(-\int_z^x \frac{g(t)}{a(t)} dt\right), \quad x \in (z, x_F),$$

where  $c$  and  $g$  are measurable functions satisfying  $c(x) \rightarrow c > 0$ ,  $g(x) \rightarrow 1$  for  $x \uparrow x_F$  and  $a(x) > 0$  with density  $a'$  satisfying  $\lim_{x \uparrow x_F} a'(x) = 0$ .

If  $F \in \text{MDA}(H_0)$ , the normalizing sequences can be chosen as  $c_n = a(d_n)$  for  $a(x) = \int_x^{x_F} \bar{F}(t) dt / \bar{F}(x)$ ,  $x < x_F$ , (the **mean excess function**), and  $d_n = F^{\leftarrow}(1 - 1/n)$ ,  $n \in \mathbb{N}$ .

## Derivation of the Hill estimator

Let  $e$  be the mean excess function for  $\log X$ . Using partial integration ( $\int H dG = [HG] - \int G dH$ ), we obtain

$$\begin{aligned} e(\log u) &= \mathbb{E}(\log X - \log u \mid \log X > \log u) \\ &= \frac{1}{\bar{F}(u)} \int_u^\infty (\log x - \log u) dF(x) = -\frac{1}{\bar{F}(u)} \int_u^\infty \log\left(\frac{x}{u}\right) d\bar{F}(x) \\ &= -\frac{1}{\bar{F}(u)} \left( \underbrace{\left[ \log\left(\frac{x}{u}\right) \bar{F}(x) \right]_u^\infty}_{=0} - \int_u^\infty \bar{F}(x) \frac{1}{x} dx \right) \\ &= \frac{1}{\bar{F}(u)} \int_u^\infty \frac{\bar{F}(x)}{x} dx = \frac{1}{\bar{F}(u)} \int_u^\infty x^{-\alpha-1} L(x) dx. \end{aligned}$$

For  $u$  sufficiently large,  $L(x) \approx L(u)$ ,  $x \geq u$  (by **Karamata's Theorem**), so

$$e(\log u) \underset{u \text{ large}}{\approx} \frac{L(u)u^{-\alpha}/\alpha}{\bar{F}(u)} = \frac{1}{\alpha}.$$

For  $n$  large and  $k$  sufficiently small, replace  $e(\cdot)$  by  $e_n(\cdot)$  and use  $u = X_{k,n}$ .

We obtain that

$$\begin{aligned}\frac{1}{\alpha} &\approx e_n(\log X_{k,n}) = \frac{\sum_{i=1}^n (\log X_i - \log X_{k,n}) I_{\{\log X_i > \log X_{k,n}\}}}{\sum_{i=1}^n I_{\{\log X_i > \log X_{k,n}\}}} \\ &= \frac{\sum_{i=1}^{k-1} (\log X_{i,n} - \log X_{k,n})}{k-1} = \frac{1}{k-1} \sum_{i=1}^{k-1} \log X_{i,n} - \log X_{k,n}\end{aligned}$$

The standard form of the estimator is typically written with the average taken over the largest  $k$  (instead of  $k-1$ ) terms.



## Non-iid data

- If  $X_1, \dots, X_n$  are serially dependent and show no tendency of clusters of extreme values (extremal index  $\theta = 1$ ), asymptotic theory of point processes suggests a limiting model for high-level threshold exceedances, in which exceedances occur according to a Poisson process and the excess losses are iid generalized Pareto distributed.
- If extremal clustering is present ( $\theta < 1$ ; e.g. GARCH processes), the assumption of independent excess losses is less satisfactory. Easiest approach: neglect the problem, simply apply MLE which is then a quasi-MLE (QMLE) (likelihood misspecified); point estimates should still be reasonable, standard errors may be too small.
- See the following section for more details on threshold exceedances.

## Point process models

So far: **loss size distribution**. Now: **loss frequency distribution**

### Threshold exceedances for strict white noise

- Consider a **strict white noise**  $(X_i)_{i \in \mathbb{N}}$  (iid from  $F \in \text{MDA}(H_\xi)$ ; can be extended to dependent processes with extremal index  $\theta = 1$ ).
- Let  $u_n(x) = c_n x + d_n$  ( $x$  fixed). We know  $F^n(u_n(x)) \xrightarrow{n \uparrow \infty} H_\xi(x)$ .  
Taking  $-\log(\cdot)$  and using  $-\log y \approx 1 - y$  for  $y \rightarrow 1$ , we obtain  
 $n\bar{F}(u_n(x)) \approx -n \log F(u_n(x)) = -\log(F^n(u_n(x))) \xrightarrow{n \uparrow \infty} -\log H_\xi(x)$ .
- $N_{u_n(x)}$  (exceedances among  $X_1, \dots, X_n$ ) fulfills  $N_{u_n(x)} \sim B(n, \bar{F}(u_n(x)))$
- The **Poisson Limit Theorem** ( $n \rightarrow \infty$ ,  $p = \bar{F}(u_n(x)) \rightarrow 0$ ,  $np = n\bar{F}(u_n(x)) \rightarrow \lambda = -\log H_\xi(x)$ ) **implies**  $N_{u_n(x)} \xrightarrow{n \uparrow \infty} \text{Poi}(-\log H_\xi(x))$ .
- One can show: **Not only is**  $N_{u_n(x)}$  **asymptotically Poisson**, but the **exceedances occur according to a Poisson process**.

## 1) On point processes

- Suppose  $Y_1, \dots, Y_n$  take values in some *state space*  $\mathcal{X}$  (e.g.  $\mathbb{R}$ ,  $\mathbb{R}^2$ ). Define for any  $A \subseteq \mathcal{X}$ , the counting rv

$$N(A) = \sum_{i=1}^n I_{\{Y_i \in A\}}.$$

Under technical conditions, see Embrechts, Klüppelberg, et al. (1997, pp. 220),  $N(\cdot)$  defines a point process.

- $N(\cdot)$  is a *Poisson point process* on  $\mathcal{X}$  with *intensity measure*  $\Lambda$  if:

1) For  $A \subseteq \mathcal{X}$  and  $k \geq 0$ ,

$$\mathbb{P}(N(A) = k) = \begin{cases} e^{-\Lambda(A)} \frac{\Lambda(A)^k}{k!}, & \text{if } \Lambda(A) < \infty, \\ 0, & \text{if } \Lambda(A) = \infty. \end{cases}$$

2)  $N(A_1), \dots, N(A_m)$  are independent for any mutually disjoint subsets  $A_1, \dots, A_m$  of  $\mathcal{X}$ .

- Note that  $\mathbb{E}N(A) = \Lambda(A)$ . Also, the *intensity (function)* is the function  $\lambda(\mathbf{x})$  which satisfies  $\Lambda(A) = \int_A \lambda(\mathbf{x}) d\mathbf{x}$ .

## 2) Asymptotic behaviour of the point process of exceedances

- For  $n \in \mathbb{N}$  and  $i \in \{1, \dots, n\}$  let  $Y_{i,n} = \frac{i}{n} I_{\{X_i > u_n(x)\}}$ . The point process of exceedances over  $u_n$  is the process  $N_n(\cdot)$  with state space  $\mathcal{X} = (0, 1]$  given by

$$N_n(A) = \sum_{i=1}^n I_{\{Y_{i,n} \in A\}}, \quad A \subseteq \mathcal{X}.$$

- $N_n$  is an element of the sequence of point processes  $(N_n)$ .  $N_n$  counts the exceedances with time of occurrence in  $A$  and we are interested in the behaviour of  $N_n$  as  $n \rightarrow \infty$ .
- Embrechts, Klüppelberg, et al. (1997, Theorem 5.3.2) show that  $N_n(\cdot)$  converges in distribution on  $\mathcal{X}$  to a Poisson process  $N(\cdot)$  with intensity  $\Lambda(\cdot)$  satisfying  $\Lambda(A) = (t_2 - t_1)\lambda(x)$  for  $A = (t_1, t_2) \subseteq \mathcal{X}$ ,  $\lambda(x) = -\log H_\xi(x)$ .
- In particular,  $\mathbb{E}N_n(A) \xrightarrow{n \uparrow \infty} \mathbb{E}N(A) = \Lambda(A) = (t_2 - t_1)\lambda(x)$ .  $\lambda$  does not depend on time and takes the constant value  $\lambda = \lambda(x)$ .

- We refer to the limiting process as a *homogeneous Poisson process with intensity* (or rate)  $\lambda$ .

### 3) Application of the result in practice

- Fix a large  $n$  and  $u = c_n x + d_n$  for some  $x$ .
- Approximate  $N_u$  by a Poisson rv and the point process of exceedances of  $u$  by a homogeneous Poisson process with rate  $\lambda = -\log H_\xi(x) = -\log H_\xi((u - d_n)/c_n) = -\log H_{\xi, \mu=d_n, \sigma=c_n}(u)$ .  
 $\Rightarrow$  Relationship between the GEV model and a Poisson model for the occurrence in time of exceedances of  $u$ .
- We see that exceedances of iid data over  $u$  are separated by iid exponential waiting times.

## The POT model

- Putting the pieces together, we obtain an asymptotic model for threshold exceedances in regularly spaced iid data (or data with  $\theta = 1$ ).
- This so-called *peaks-over-threshold (POT) model* makes the following assumptions:
  - 1) Exceedance times occur according to a homogeneous Poisson process.
  - 2) Excesses above  $u$  are iid and independent of exceedance times.
  - 3) The excess distribution is generalized Pareto.
- This model can also be viewed as a *marked Poisson point process* (exceedance times = points; GPD-distributed excesses = marks) or a (non-homogeneous) *two-dimensional Poisson* point process (point  $(t, x)$  = (time, magnitude of exceedance))

## 1) Two-dimensional Poisson formulation of POT model

- Assume that, on the state space  $\mathcal{X} = (0, 1] \times (u, \infty)$ , the point process defined by  $N(A) = \sum_{i=1}^n I_{\{(i/n, X_i) \in A\}}$  is a Poisson process with intensity at  $(t, x)$  given by

$$\lambda(x) = \lambda(t, x) = \begin{cases} \frac{1}{\sigma} (1 + \xi \frac{x-\mu}{\sigma})^{-1/\xi-1}, & \text{if } (1 + \xi(x - \mu)/\sigma) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

- For  $A = (t_1, t_2) \times (x, \infty) \subseteq \mathcal{X}$ , the intensity measure is

$$\Lambda(A) = \int_{t_1}^{t_2} \int_x^{\infty} \lambda(y) dy dt = -(t_2 - t_1) \log H_{\xi, \mu, \sigma}(x)$$

Thus, for any  $x \geq u$ , the one-dimensional process of exceedances of  $x$  is a homogeneous Poisson process with intensity  $\tau(x) = -\log H_{\xi, \mu, \sigma}(x)$ .

- $\bar{F}_u(x)$  can be calculated as the ratio of the rates of exceeding  $u + x$  and  $u$  via

$$\bar{F}_u(x) = \frac{\tau(u+x)}{\tau(u)} = \left(1 + \frac{\xi x}{\sigma + \xi(u - \mu)}\right)^{-1/\xi} = \bar{G}_{\xi, \sigma + \xi(u - \mu)}(x)$$

This is precisely the **POT model**.

- The model also implies the **GEV model**. Consider  $\{M_n \leq x\}$  for some  $x \geq u$ , i.e. the event that there are no points in  $A = (0, 1] \times (x, \infty)$ . Thus,  $\mathbb{P}(M_n \leq x) = \mathbb{P}(N(A) = 0) = \exp(-\Lambda(A)) = H_{\xi, \mu, \sigma}(x)$ ,  $x \geq u$ , which is precisely the GEV model.

## 2) Statistical estimation of the POT model

- Given the **exceedances**  $\tilde{X}_1 < \dots < \tilde{X}_{N_u}$ ,  $A = (0, 1] \times (u, \infty)$  and  $\Lambda(A) = \tau(u) =: \tau_u$ , the **likelihood**  $L(\xi, \sigma, \mu; \tilde{X}_1, \dots, \tilde{X}_{N_u})$  is

$$\underbrace{N_u!}_{\text{ordered sample}} \underbrace{e^{-\Lambda(A)} \frac{\Lambda(A)^{N_u}}{N_u!}}_{\text{prob. of } N_u \text{ samples}} \underbrace{\prod_{i=1}^{N_u} \frac{\lambda(\tilde{X}_i)}{\Lambda(A)}}_{\text{density of } \tilde{X}_i} = e^{-\Lambda(A)} \prod_{i=1}^{N_u} \lambda(\tilde{X}_i) = e^{-\tau_u} \prod_{i=1}^{N_u} \lambda(\tilde{X}_i).$$

- Reparametrizing  $\lambda$  by  $\tau_u = -\log H_{\xi, \mu, \sigma}(u) = (1 + \xi \frac{u - \mu}{\sigma})^{-1/\xi}$  and



$\beta = \sigma + \xi(u - \mu)$ , we obtain

$$\begin{aligned}
\lambda(x) &= \frac{1}{\sigma} \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}-1} = \frac{1}{\sigma} \left( \left(1 + \xi \frac{u - \mu}{\sigma}\right) \left(1 + \frac{\xi \frac{x-u}{\sigma}}{1 + \xi \frac{u-\mu}{\sigma}}\right) \right)^{-\frac{1}{\xi}-1} \\
&= \frac{\tau_u}{\sigma(1 + \xi \frac{u-\mu}{\sigma})} \left(1 + \frac{\xi \frac{x-u}{\sigma}}{1 + \xi \frac{u-\mu}{\sigma}}\right)^{-\frac{1}{\xi}-1} = \frac{\tau_u}{\beta} \left(1 + \frac{\xi(x-u)}{\sigma + \xi(u-\mu)}\right)^{-\frac{1}{\xi}-1} \\
&= \frac{\tau_u}{\beta} \left(1 + \frac{\xi(x-u)}{\beta}\right)^{-\frac{1}{\xi}-1} = \tau_u g_{\xi, \beta}(x-u),
\end{aligned}$$

where  $\xi \in \mathbb{R}$  and  $\tau_u, \beta > 0$ . Therefore,  $\ell(\xi, \sigma, \mu; \tilde{X}_1, \dots, \tilde{X}_{N_u})$  equals

$$\begin{aligned}
&= -\tau_u + \sum_{i=1}^{N_u} \log \lambda(\tilde{X}_i) = -\tau_u + N_u \log \tau_u + \sum_{i=1}^{N_u} \overbrace{(\log \lambda(\tilde{X}_i) - \log \tau_u)}^{= \log g_{\xi, \beta}(\tilde{X}_i - u)} \\
&= \ell_{\text{Poi}}(\tau_u; N_u) - N_u \log(T) + \ell_{\text{GPD}}(\xi, \beta; \tilde{X}_1 - u, \dots, \tilde{X}_{N_u} - u),
\end{aligned} \tag{135}$$

where  $\ell_{\text{Poi}}$  is the log-likelihood for a one-dimensional homogeneous Poisson process with rate  $\tau_u$  and  $\ell_{\text{GPD}}$  is the log-likelihood for fitting a GPD to the excesses  $\tilde{X}_i - u$ ,  $i \in \{1, \dots, N_u\}$ .

- We can thus separate inferences about  $(\xi, \beta)$  and  $\tau_u$ . Estimate  $(\xi, \beta)$  in a GPD analysis and then  $\tau_u$  by its MLE  $N_u$ . Use these estimates to infer estimates of  $\mu = u - \beta(1 - \tau_u^\xi)/\xi$  and  $\sigma = \tau_u^\xi \beta$ .

### 3) Advantages of the POT model formulation

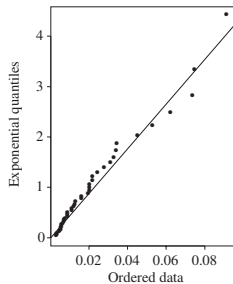
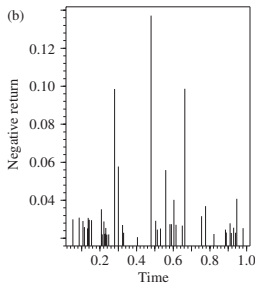
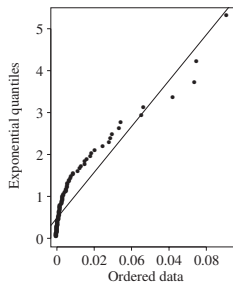
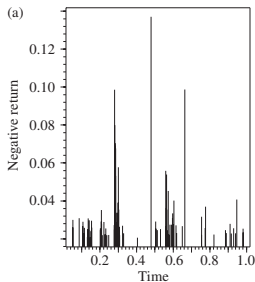
- One advantage of the two-dimensional Poisson point process model is that  $\xi$ ,  $\mu$  and  $\sigma$  do not depend on  $u$  (unlike  $\beta$  in the GPD model).  
⇒ In practice, we would expect the estimated parameters of the Poisson model to be roughly stable over a range of high thresholds.
- The intensity  $\lambda$  is thus often used to introduce covariates to obtain Poisson processes which are non-homogeneous in time, e.g. by replacing  $\mu$  and  $\sigma$  by parameters that vary over time as functions of covariates; see, e.g. Chavez-Demoulin et al. (2014).

#### 4) Applicability of the POT model to return series data

- Returns do not really form genuine point events in time (in contrast to, e.g. water levels). They are discrete-time measurements that describe short-term changes (a day or a week). Nonetheless, assume that under a longer-term perspective, such data can be approximated by point events in time.
- Exceedances of  $u$  for daily financial return series do not necessarily occur according to a homogeneous Poisson process. They tend to cluster. Thus the standard POT model is not directly applicable.
- For stochastic processes with extremal index  $\theta < 1$ , e.g. GARCH processes, the extremal clusters themselves should occur according to a homogeneous Poisson process in time  $\Rightarrow$  Individual exceedances occur according to a *Poisson cluster process*; see Leadbetter (1991). Thus a suitable model for the occurrence and magnitude of exceedances in a financial return series might be some form of marked Poisson cluster

process.

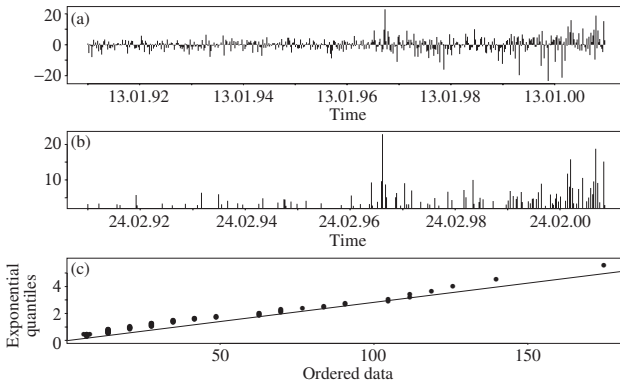
- *Declustering* may circumvent the problem. One identifies clusters (ad hoc; not easy) of exceedances and then applies the POT model to cluster maxima only.
- A possible declustering algorithm is the *runs method*. A run size  $r$  is fixed and two successive exceedances are said to belong to two different clusters if they are separated by a run of at least  $r$  values below  $u$ ; see Embrechts, Klüppelberg, et al. (1997, pp. 422).
- In the following figure the DAX daily negative returns have been declustered with  $r = 10$  trading days; this reduces the 100 exceedances to 42 cluster maxima.



- (a): DAX daily negative returns and a Q-Q plot of their spacings
- (b): Declustered data (runs method with  $r = 10$  trading days  $\Rightarrow$  spacings are more consistent with a Poisson model)
- However, by neglecting the modelling of cluster formation, we cannot make more dynamic statements about the intensity of occurrence of exceedances.

## Example A.10 (POT analysis of AT&T weekly losses (continued))

Consider the 102 weekly percentage losses exceeding  $u = 2.75\%$ :



- Inter-exceedance times seem to follow an exponential distribution.
- But exceedances become more frequent over time (which contradicts a homogeneous Poisson process)

- Using the log-likelihood (135), we fit a two-dimensional Poisson model to the 102 exceedances of  $u = 2.75\%$ . The parameter estimates are  $\hat{\xi} = 0.22$ ,  $\hat{\mu} = 19.9$  and  $\hat{\sigma} = 5.95$ .
- The implied GPD scale parameter is  $\hat{\beta} = \hat{\sigma} + \hat{\xi}(u - \hat{\mu}) = 2.1 \Rightarrow$  The same  $\hat{\xi}$  and  $\hat{\beta}$  as in Example 5.19.
- The estimated exceedance rate over  $u = 2.75$  is  $\hat{\tau}(u) = -\log H_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}(u) = 102$  (= number of exceedances; as theory suggests).
- Higher thresholds, e.g. 15%: Since  $\hat{\tau}(15) = 2.50$ , losses exceeding 15% occur as a Poisson process with rate 2.5 losses per 10-year period ( $\approx$  a four-year event).  $\Rightarrow$  The Poisson model provides an alternative method of defining the return period of an event.
- Similarly, estimate return levels: If the 10-year return level is the level which is exceeded according to a Poisson process with rate one loss per 10 years, estimate the level by solving  $\hat{\tau}(u) = 1$  w.r.t.  $u$ , so  $u = H_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}^{-1}(\exp(-1)) = 19.9$  so the 10-year event is a weekly loss

of roughly 20%.

- Confidence intervals for such quantities can be constructed via profile likelihoods.



## A.6 Multivariate models

### Conditional distributions and independence

*Proof of (16).* We have

$$\begin{aligned}& \int_{(-\infty, x_1]} F_{X_2|X_1}(x_2 | z) dF_{X_1}(z) \\&= \int_{\mathbb{R}^d} I_{\{z \leq x_1\}} \mathbb{E}(I_{\{X_2 \leq x_2\}} | \mathbf{X}_1 = z) dF_{X_1}(z) \\&= \mathbb{E}(I_{\{\mathbf{X}_1 \leq x_1\}} \mathbb{E}(I_{\{X_2 \leq x_2\}} | \mathbf{X}_1)) = \mathbb{E}(\mathbb{E}(I_{\{\mathbf{X}_1 \leq x_1, X_2 \leq x_2\}} | \mathbf{X}_1)) \\&\stackrel{\text{tower property}}{=} \mathbb{E}(I_{\{\mathbf{X}_1 \leq x_1, X_2 \leq x_2\}}) = F_{X_1, X_2}(x_1, x_2),\end{aligned}$$

where the second-last equality holds by the **tower property** of conditional expectations. □

## The multivariate normal distribution

*Proof of the form of the cf of  $N(0, 1)$ ; see Proposition 6.3.* The rv  $Z \sim N(0, 1)$  has density  $\varphi(x) = \exp(-x^2/2)/\sqrt{2\pi}$  which satisfies

- i)  $\varphi(x) = \varphi(-x)$ ;
- ii)  $\varphi'(x) = -x\varphi(x)$ .

By Euler's Formula, the characteristic function  $\phi_Z(t)$  of  $Z$  is given by

$$\phi_Z(t) = \int_{-\infty}^{\infty} (\cos(tx) + i \sin(tx))\varphi(x) dx \stackrel{\text{i)}}{=} \int_{-\infty}^{\infty} \cos(tx)\varphi(x) dx.$$

Hence,

$$\phi'_Z(t) = \int_{-\infty}^{\infty} \sin(tx)(-x)\varphi(x) dx \stackrel{\text{ii)}}{=} \int_{-\infty}^{\infty} \sin(tx)\varphi'(x) dx \stackrel{\text{by parts}}{=} -t\phi_Z(t).$$

We also know that  $\phi_Z(0) = 1$ . This initial value problem has the unique solution  $\phi_Z(t) = \exp(-t^2/2)$ . □

### Theorem A.11 (Cramér–Wold)

Let  $\mathbf{X}, \mathbf{X}_n$ ,  $n \in \mathbb{N}$ , be random vectors. Then

$$\mathbf{X}_n \xrightarrow[n \uparrow \infty]{d} \mathbf{X} \iff \mathbf{a}'\mathbf{X}_n \xrightarrow[n \uparrow \infty]{d} \mathbf{a}'\mathbf{X} \quad \forall \mathbf{a} \in \mathbb{R}^d$$

*Proof.*

“ $\Rightarrow$ ” This follows from the Continuous Mapping Theorem with the map  $g(\mathbf{x}) = \mathbf{a}'\mathbf{x}$ .

“ $\Leftarrow$ ” Note that  $\phi_{\mathbf{X}_n}(\mathbf{t}) = \mathbb{E}(\exp(i \cdot \mathbf{1} \cdot \mathbf{t}'\mathbf{X}_n)) = \phi_{\mathbf{t}'\mathbf{X}_n}(1) \xrightarrow[n \uparrow \infty]{} \phi_{\mathbf{t}'\mathbf{X}}(1) = \phi_{\mathbf{X}}(\mathbf{t})$  for all  $\mathbf{t}$  and apply the Lévy Continuity Theorem.  $\square$

### Corollary A.12

Let  $\mathbf{X}, \mathbf{Y}$  be two random vectors. Then

$$\mathbf{X} \stackrel{d}{=} \mathbf{Y} \iff \mathbf{a}'\mathbf{X} \stackrel{d}{=} \mathbf{a}'\mathbf{Y} \quad \forall \mathbf{a} \in \mathbb{R}^d.$$

## Properties of multivariate normal variance mixtures

*Proof of Lemma 6.10.* W.l.o.g. assume  $\mu = \mathbf{0}$ .

$$\begin{aligned} \Rightarrow \mathbb{E}|X_i| \mathbb{E}|X_j| &\stackrel{\text{ind.}}{=} \mathbb{E}(|X_i||X_j|) = \mathbb{E}(W|Z_i||Z_j|) \stackrel{\text{ind.}}{=} \mathbb{E}(W) \mathbb{E}|Z_i| \mathbb{E}|Z_j| \\ &\stackrel{\text{Jensen}}{\geq} \mathbb{E}(\sqrt{W})^2 \mathbb{E}|Z_i| \mathbb{E}|Z_j| \stackrel{\text{ind.}}{=} \mathbb{E}|\sqrt{W}Z_i| \mathbb{E}|\sqrt{W}Z_j| = \mathbb{E}|X_i| \mathbb{E}|X_j| \end{aligned}$$

$\Rightarrow$  We must have “=” in Jensen’s inequality. This holds if and only if  $W$  is constant a.s.; so  $\mathbf{X} \sim N_d(\mathbf{0}, W I_d)$  in this case.

$\Leftarrow$   $W$  a.s. constant  $\Rightarrow \mathbf{X} \sim N_d(\mathbf{0}, W I_d) \Rightarrow X_i, X_j$  independent.  $\square$

## Spherical distributions

*Proof of Theorem 6.15.*

1)  $\Rightarrow$  2):  $\phi_{\mathbf{Y}}(\mathbf{t}) = \phi_{U\mathbf{Y}}(\mathbf{t}) = \phi_{\mathbf{Y}}(U'\mathbf{t})$  for all  $U \in \mathbb{R}^{d \times d}$  orthogonal. Since  $U$  can only change the direction of  $\mathbf{t}$  but not its length,  $\phi_{\mathbf{Y}}(\mathbf{t})$  only depends on  $\|\mathbf{t}\|$ , i.e. the length of  $\mathbf{t} \Rightarrow$  we can define  $\psi(\|\mathbf{t}\|^2) = \phi_{\mathbf{Y}}(\mathbf{t})$ .

$$\begin{aligned}
2) \Rightarrow 3): \phi_{Y_1}(t) &= \phi_Y(te_1) \stackrel{2)}{=} \psi(t^2) \quad (*). \quad \text{Now } \phi_{a'Y}(t) = \phi_Y(ta) \stackrel{2)}{=} \\
&\psi(t^2 \|a\|^2) = \psi((t\|a\|)^2) \stackrel{(*)}{=} \phi_{Y_1}(t\|a\|) = \phi_{\|a\|Y_1}(t) \\
3) \Rightarrow 1): \phi_{UY}(t) &= \mathbb{E}(\exp(i(U't)'Y)) \stackrel{U't:=a}{=} \mathbb{E}(\exp(ia'Y)) \stackrel{3)}{=} \mathbb{E}(\exp(i\|a\|Y_1)) \\
&= \mathbb{E}(\exp(i\|t\|Y_1)) \stackrel{3)}{=} \mathbb{E}(\exp(it'Y)) = \phi_Y(t) \quad \square
\end{aligned}$$

*Proof of Theorem 6.16.* Let  $\Omega_d$  be the characteristic generator of  $S$ .

$$\begin{aligned}
\text{"}\Rightarrow\text{" } Y &\sim S_d(\psi) \Rightarrow \phi_Y(\|t\|u) \stackrel{2)}{=} \psi(\|t\|^2 u'u) = \psi(\|t\|^2) \text{ for all } u \in \mathbb{R}^d : \\
&\|u\| = 1. \text{ Replacing } u \text{ by } S \text{ and integrating leads to } \psi(\|t\|^2) = \\
&\mathbb{E}_S(\phi_Y(\|t\|S)) = \mathbb{E}_S(\mathbb{E}_Y(e^{i\|t\|S'Y})) \stackrel{\text{Fubini}}{=} \mathbb{E}_Y(\mathbb{E}_S(e^{i\|t\|S'Y})) = \\
&\mathbb{E}_Y(\phi_S(\|t\|Y)) \stackrel{2)}{=} \mathbb{E}_Y(\Omega_d(\|t\|^2 Y'Y)). \text{ We thus obtain that} \\
\phi_Y(t) &\stackrel{2)}{=} \psi(\|t\|^2) \stackrel{R:=\|Y\|}{=} \mathbb{E}_R(\Omega_d(\|t\|^2 R^2)) = \int_0^\infty \Omega_d(\|t\|^2 r^2) dF_R(r) \\
&= \int_0^\infty \phi_S(rt) dF_R(r) = \phi_{RS}(t) \text{ for all } t \in \mathbb{R}^d.
\end{aligned}$$

$\Leftarrow$  Let  $Z \sim N_d(0, I_d)$ . Since  $Z$  is spherical and  $\|Z/\|Z\|\| = \|Z\|/\|Z\| =$

1,  $\mathbf{S} \stackrel{d}{=} \mathbf{Z}/\|\mathbf{Z}\|$ . As such,  $\mathbf{S}$  itself is spherical, since  $U\mathbf{S} \stackrel{d}{=} U\mathbf{Z}/\|\mathbf{Z}\| \stackrel{d}{=} \mathbf{Z}/\|\mathbf{Z}\| \stackrel{d}{=} \mathbf{S}$  for any orthogonal  $U \in \mathbb{R}^{d \times d}$ . Theorem 6.15 Part 2) implies that  $\phi_{\mathbf{S}}(\mathbf{t}) = \Omega_d(\|\mathbf{t}\|^2)$ , so  $\phi_{R\mathbf{S}}(\mathbf{t}) = \mathbb{E}(\exp(i\mathbf{t}'R\mathbf{S})) = \mathbb{E}_R(\mathbb{E}(\exp(i\mathbf{t}'R\mathbf{S}) | R)) = \mathbb{E}_R(\phi_{\mathbf{S}}(R\mathbf{t})) = \mathbb{E}_R(\Omega_d(R^2\|\mathbf{t}\|^2))$ , which is a function of  $\|\mathbf{t}\|^2$  and thus, by 2),  $R\mathbf{S}$  is spherical.  $\square$

### Density of $\mathbf{Y} \sim S_d(\psi)$ constant on spheres

If  $\mathbf{Y}$  admits a density  $f_{\mathbf{Y}}$ , then  $f_{\mathbf{Y}}(\mathbf{y})$  is constant on hyperspheres in  $\mathbb{R}^d$ . The *inversion formula*  $f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\mathbf{t}'\mathbf{y}} \phi_{\mathbf{Y}}(\mathbf{t}) d\mathbf{t}$  and Theorem 6.15 Part 2) show that for any orthogonal  $U$ ,

$$\begin{aligned} f_{\mathbf{Y}}(U\mathbf{y}) &\stackrel{\text{inv.}}{=} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(U'\mathbf{t})'\mathbf{y}} \phi_{\mathbf{Y}}(\mathbf{t}) d\mathbf{t} \\ &\stackrel{\text{subs.}}{=} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\mathbf{s}'\mathbf{y}} \phi_{\mathbf{Y}}(U\mathbf{s}) d\mathbf{s} \\ &\stackrel{2)}{=} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\mathbf{s}'\mathbf{y}} \psi((U\mathbf{s})'U\mathbf{s}) d\mathbf{s} \end{aligned}$$

$$= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-is'y} \psi(s's) ds \underset{\text{backwards}}{=} \cdots = f_Y(y).$$

This implies that  $f_Y(y) = g(\|y\|^2)$  for a function  $g : [0, \infty) \rightarrow [0, \infty)$ , the density generator. For  $Y \sim t_d(\nu, \mathbf{0}, I_d)$ , one can show that  $g(x) = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)(\pi\nu)^{d/2}} (1 + \frac{x}{\nu})^{-(\nu+d)/2}$ .

## Elliptical distributions

### Proposition A.13

Let  $X \sim E_d(\mu, \Sigma, \psi)$  for positive definite  $\Sigma$  and  $\mathbb{E}(R^2) < \infty$  (i.e.  $\text{cov}(X)$  finite). For any  $c \geq 0$  such that  $\mathbb{P}((X - \mu)' \Sigma^{-1} (X - \mu) \geq c) > 0$ ,

$$\text{corr}(X \mid (X - \mu)' \Sigma^{-1} (X - \mu) \geq c) = \text{corr}(X).$$

*Proof.*  $X \mid ((X - \mu)' \Sigma^{-1} (X - \mu) \geq c) \stackrel{d}{=} \mu + RAS \mid (R^2 \geq c) \stackrel{\text{ind.}}{=} \mu + \tilde{R}AS$   
 where  $\tilde{R} \stackrel{d}{=} (R \mid R^2 \geq c)$ .  $A$  (and thus  $\Sigma$ ) remains the same.  $\square$

## Estimating scale and correlation

- Suppose  $\mathbf{X}_1, \dots, \mathbf{X}_n \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ . How can we estimate  $\boldsymbol{\mu}, \Sigma$  and  $P$ ? ( $P$  is the correlation matrix corresponding to  $\Sigma$ ; this always exists)
- $\bar{\mathbf{X}}, S, R$  may not be the best options for heavy-tailed data (e.g. concerning robustness against contamination).

## M-estimators for $\boldsymbol{\mu}, \Sigma$ (see Maronna (1976))

- **Goal:** Improve given estimators  $\hat{\boldsymbol{\mu}}, \hat{\Sigma}$ .
- **Idea:** Compute improved estimates by downweighting observations with large  $D_i = \sqrt{(\mathbf{X}_i - \hat{\boldsymbol{\mu}})' \hat{\Sigma}^{-1} (\mathbf{X}_i - \hat{\boldsymbol{\mu}})}$  (these are the ones which tend to distort  $\hat{\boldsymbol{\mu}}, \hat{\Sigma}$  most).
- This can be turned into an iterative procedure that converges to so-called *M-estimates* of location and scale ( $\hat{\Sigma}$  is in general biased).



### Algorithm A.14 (M-estimators of location and scale)

1) Set  $k = 1$ ,  $\hat{\mu}^{[1]} = \bar{X}$  and  $\hat{\Sigma}^{[1]} = S$ .

2) Repeat until convergence:

2.1) For  $i \in \{1, \dots, n\}$  set  $D_i = \sqrt{(\mathbf{X}_i - \hat{\mu}^{[k]})' \hat{\Sigma}^{[k]-1} (\mathbf{X}_i - \hat{\mu}^{[k]})}$ .

2.2) Update:

$$\hat{\mu}^{[k+1]} = \frac{\sum_{i=1}^n w_1(D_i) \mathbf{X}_i}{\sum_{i=1}^n w_1(D_i)},$$

where  $w_1$  is a weight function, e.g.  $w_1(x) = (d + \nu)/(x^2 + \nu)$  (or  $I_{x \leq a} + (a/x)I_{x > a}$  for some value  $a$ ).

2.3) Update:

$$\hat{\Sigma}^{[k+1]} = \frac{1}{n} \sum_{i=1}^n w_2(D_i^2) (\mathbf{X}_i - \hat{\mu}^{[k]})(\mathbf{X}_i - \hat{\mu}^{[k]})',$$

where  $w_2$  is a weight function, e.g.  $w_2(x) = w_1(\sqrt{x})$  (or  $(w_1(\sqrt{x}))^2$ ).

2.4) Set  $k$  to  $k + 1$ .

## Factor models

### Example A.15 (One-factor/equicorrelation model)

Let  $\mathbb{E}(\mathbf{X}) = \mathbf{0}$ ,  $\Sigma = \text{cov}(\mathbf{X}) = \rho J_d + (1 - \rho)I_d$  ( $J_d = (1) \in \mathbb{R}^{d \times d}$ ).

- Then  $\Sigma = BB' + \Upsilon$  for  $B = \sqrt{\rho}\mathbf{1}$  and  $\Upsilon = (1 - \rho)I_d$ .
- Any  $Y$  with  $\mathbb{E}Y = 0$ ,  $\text{var } Y = 1$  independent of  $\mathbf{X}$  leads to the *factor decomposition* of  $\mathbf{X}$

$$F = \frac{\sqrt{\rho}}{1 + \rho(d-1)} \sum_{j=1}^d X_j + \sqrt{\frac{1-\rho}{1 + \rho(d-1)}} Y, \quad \varepsilon_j = X_j - \sqrt{\rho}F.$$

We have  $\mathbb{E}(F) = 0$ ,  $\text{var}(F) = 1$ , so  $\mathbf{X} = \mathbf{0} + BF + \boldsymbol{\varepsilon} = \sqrt{\rho}\mathbf{1}F + \boldsymbol{\varepsilon}$ .

- The requirements of Definition 6.25 are fulfilled since  $\text{cov}(F, \varepsilon_j) = 0$ ,  $\text{cov}(\varepsilon_j, \varepsilon_k) = 0$  for all  $j \neq k$ .
- $\text{var}(\bar{X}_n) = \text{var}(\sqrt{\rho}F + \bar{\varepsilon}_d) = \rho + \frac{1-\rho}{d} \xrightarrow{(d \rightarrow \infty)} \rho$  (systematic factor matters!)

- If  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ , take  $Y \sim \mathcal{N}(0, 1)$  (then  $F$  is also normal). One typically writes this (one-factor) equicorrelation model as  $\mathbf{X} = \sqrt{\rho}F + \sqrt{1 - \rho}\mathbf{Z}$ , where  $F, Z_1, \dots, Z_d \stackrel{\text{ind.}}{\sim} \mathcal{N}(0, 1)$ .

## Multivariate regression

- Here, construct large matrices:

$$X = \underbrace{\begin{pmatrix} \mathbf{X}'_1 \\ \vdots \\ \mathbf{X}'_n \end{pmatrix}}_{n \times d}, \quad F = \underbrace{\begin{pmatrix} 1 & \mathbf{F}'_1 \\ \vdots & \vdots \\ 1 & \mathbf{F}'_n \end{pmatrix}}_{n \times (p+1)}, \quad \tilde{B} = \underbrace{\begin{pmatrix} \mathbf{a}' \\ B' \end{pmatrix}}_{(p+1) \times d}, \quad E = \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}'_1 \\ \vdots \\ \boldsymbol{\varepsilon}'_n \end{pmatrix}}_{n \times d}.$$

This model can be expressed by  $X = F\tilde{B} + E$  (estimate  $\tilde{B}$ ).

- Assume the unobserved  $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$  form a white noise process. Then, conditional on  $\mathbf{F}_1, \dots, \mathbf{F}_n$ , we have a multivariate linear regression, see, e.g. Mardia et al. (1979), with estimator  $\hat{\tilde{B}} = (F'F)^{-1}F'X$ .

- Now examine the conditions of Definition 6.25: Do the errors vectors  $\varepsilon_t$  come from a distribution with diagonal covariance matrix, and are they uncorrelated with the factors?
- Consider the sample correlation matrix of  $\hat{E} = X - F\hat{B}$  (model residual matrix; hopefully shows that there is little correlation in the errors) and take the diagonal elements as an estimator  $\hat{\Upsilon}$  of  $\Upsilon$ .

## Sample principal components

- Assume  $X_1, \dots, X_n$  with identical distribution, **unknown mean vector  $\mu$  and covariance matrix  $\Sigma$**  with the **spectral decomposition  $\Sigma = \Gamma\Lambda\Gamma'$**  as before.
- Estimate  $\mu$  by  $\bar{X}$  and  $\Sigma$  by  $S_x = \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X})(X_t - \bar{X})'$ .
- Apply the **spectral decomposition** to  $S_x$  to get  $S_x = GLG'$ , where  $G$  is the eigenvector matrix and  $L = \text{diag}(l_1, \dots, l_d)$  is the diagonal matrix consisting of ordered eigenvalues.

- Define the “sample principle component transforms”  $\mathbf{Y}_t = G'(\mathbf{X}_t - \bar{\mathbf{X}})$ ,  $t \in \{1, \dots, n\}$ . The  $j$ th component  $Y_{t,j} = \mathbf{g}'_j(\mathbf{X}_t - \bar{\mathbf{X}})$  is the  $j$ th sample principal component at time  $t$  ( $\mathbf{g}_j$  is the  $j$ th column of  $G$ ).
- The rotated vectors  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  have sample covariance matrix  $L$ :

$$\begin{aligned} S_y &= \frac{1}{n} \sum_{t=1}^n (\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_t - \bar{\mathbf{Y}})' = \frac{1}{n} \sum_{t=1}^n \mathbf{Y}_t \mathbf{Y}_t' \\ &= \frac{1}{n} \sum_{t=1}^n G'(\mathbf{X}_t - \bar{\mathbf{X}})(\mathbf{X}_t - \bar{\mathbf{X}})'G = G' S_x G = L. \end{aligned}$$

Thus the rotated vectors show no correlation between components and the components are ordered by their sample variances, from largest to smallest.

- Now use  $G$  and  $\mathbf{Y}_t$  to calibrate an approximate factor model. We assume our data are realizations from the model

$$\mathbf{X}_t = \bar{\mathbf{X}} + G_1 \mathbf{F}_t + \varepsilon_t, \quad t \in \{1, \dots, n\},$$

where  $G_1$  consists of the first  $k$  columns of  $G$  and  $F_t = (Y_{t,1}, \dots, Y_{t,k})$ ,  $t \in \{1, \dots, n\}$ .

- In practice, the errors  $\epsilon_t$  do not have a diagonal covariance matrix and are not uncorrelated with  $F_t$ . Nevertheless the method is a popular approach to constructing time series of statistically explanatory factors from multivariate time series of risk-factor changes.

## A.7 Copulas and dependence

### An example

Let  $C(\mathbf{u}) = \lambda C_1(\mathbf{u}) + (1 - \lambda)C_2(\mathbf{u})$  for copulas  $C_1, C_2$  and  $\lambda \in [0, 1]$  (convex combination). Then  $C$  is again a copula since:

#### 1) Analytical proof:

- Let  $\mathbf{u}_j = (u_1, \dots, u_{j-1}, 0, u_{j+1}, \dots, u_d)$ . Then

$$C(\mathbf{u}_j) = \lambda C_1(\mathbf{u}_j) + (1 - \lambda)C_2(\mathbf{u}_j) = \lambda \cdot 0 + (1 - \lambda) \cdot 0 = 0$$

since  $C_1, C_2$  are grounded. Hence,  $C$  is grounded.

- Let  $\mathbf{u}_j = (1, \dots, 1, u_j, 1, \dots, 1)$ . Then

$$C(\mathbf{u}_j) = \lambda C_1(\mathbf{u}_j) + (1 - \lambda)C_2(\mathbf{u}_j) = \lambda u_j + (1 - \lambda)u_j = u_j$$

since  $C_1, C_2$  have  $U[0, 1]$  margins. Hence,  $C$  has  $U[0, 1]$  margins.

- $\Delta_{(a,b]}C = \lambda \Delta_{(a,b]}C_1 + (1 - \lambda)\Delta_{(a,b]}C_2 \geq 0$ , so  $C$  is  $d$ -increasing.

## 2) Stochastic proof:

Let  $U_k \sim C_k$ ,  $k \in \{1, 2\}$  and let  $X \sim B(1, \lambda)$ , independent of each other. Furthermore, let

$$U = \begin{cases} U_1, & \text{if } X = 1, \\ U_2, & \text{if } X = 0. \end{cases}$$

The Law of Total Probability implies that

$$\begin{aligned} \mathbb{P}(U \leq \mathbf{u}) &= \mathbb{P}(U \leq \mathbf{u}, X = 1) + \mathbb{P}(U \leq \mathbf{u}, X = 0) \\ &= \mathbb{P}(U_1 \leq \mathbf{u}, X = 1) + \mathbb{P}(U_2 \leq \mathbf{u}, X = 0) \\ &= \mathbb{P}(U_1 \leq \mathbf{u}) \mathbb{P}(X = 1) + \mathbb{P}(U_2 \leq \mathbf{u}) \mathbb{P}(X = 0) \\ &= C_1(\mathbf{u}) \lambda + C_2(\mathbf{u}) (1 - \lambda) = C(\mathbf{u}). \end{aligned}$$

So  $U \sim C$  and hence  $C$  is a df. From the same calculation it follows that  $U$  has uniform margins, hence  $C$  is a copula.



## Basic properties

### Lemma A.16

For any copula  $C$ ,  $|C(\mathbf{b}) - C(\mathbf{a})| \leq \sum_{j=1}^d |b_j - a_j|$  for all  $\mathbf{a}, \mathbf{b} \in [0, 1]^d$ .

*Proof.* Using a telescoping sum expansion and the triangle inequality, we obtain

$$|C(\mathbf{b}) - C(\mathbf{a})| \leq \sum_{j=1}^d |C(b_1, \dots, b_{j-1}, b_j, a_{j+1}, \dots, a_d) - C(b_1, \dots, b_{j-1}, a_j, a_{j+1}, \dots, a_d)|.$$

W.l.o.g. let  $\mathbf{a} \leq \mathbf{b}$ . By  $d$ -increasingness,  $C \nearrow$  component-wise, so omit  $|\cdot|$ . Since, by  $d$ -increasingness, the  $j$ th summand is  $\nearrow$  in each component  $\neq j$ , let  $b_1, \dots, b_{j-1}, a_{j+1}, \dots, a_d \nearrow 1$  to obtain the upper bound  $\sum_{j=1}^d C(1, \dots, 1, b_j, 1, \dots, 1) - C(1, \dots, 1, a_j, 1, \dots, 1) = b_j - a_j$  for summand  $j$ .  $\square$

## Generalized inverses

$T \nearrow$  means that  $T$  is *increasing*;  $T \uparrow$  means that  $T$  is *strictly increasing*;  
 $\text{ran } T = \{T(x) : x \in \mathbb{R}\}$  denotes the *range* of  $T$ .

### Proposition A.17 (Working with generalized inverses)

Let  $T : \mathbb{R} \rightarrow \mathbb{R} \nearrow$  with  $T(-\infty) = \lim_{x \downarrow -\infty} T(x)$  and  $T(\infty) = \lim_{x \uparrow \infty} T(x)$   
and let  $x, y \in \mathbb{R}$ . Then,

- (GI1)  $T^{\leftarrow}(y) = -\infty$  if and only if  $T(x) \geq y$  for all  $x \in \mathbb{R}$ . Similarly,  
 $T^{\leftarrow}(y) = \infty$  if and only if  $T(x) < y$  for all  $x \in \mathbb{R}$ .
- (GI2)  $T^{\leftarrow} \nearrow$ . If  $T^{\leftarrow}(y) \in (-\infty, \infty)$ ,  $T^{\leftarrow}$  is left-continuous at  $y$  and  
admits a limit from the right at  $y$ .
- (GI3)  $T^{\leftarrow}(T(x)) \leq x$ . If  $T \uparrow$ , then  $T^{\leftarrow}(T(x)) = x$ .
- (GI4) Let  $T$  be right-continuous and  $\text{ran } T$  denote the *range of  $T$* , i.e.  
 $\text{ran } T = \{T(x) : x \in \mathbb{R}\}$ .  $T^{\leftarrow}(y) < \infty$  implies  $T(T^{\leftarrow}(y)) \geq y$ .  
Furthermore,  $y \in \text{ran } T \cup \{\inf T, \sup T\}$  implies  $T(T^{\leftarrow}(y)) = y$ .

Moreover, if  $y < \inf T$  then  $T(T^{\leftarrow}(y)) > y$  and if  $y > \sup T$  then  $T(T^{\leftarrow}(y)) < y$ .

- (GI5)  $T(x) \geq y$  implies  $x \geq T^{\leftarrow}(y)$ . The other implication holds if  $T$  is right-continuous. Furthermore,  $T(x) < y$  implies  $x \leq T^{\leftarrow}(y)$ .
- (GI6)  $(T^{\leftarrow}(y-), T^{\leftarrow}(y+)) \subseteq \{x \in \mathbb{R} : T(x) = y\} \subseteq [T^{\leftarrow}(y-), T^{\leftarrow}(y+)]$ , where  $T^{\leftarrow}(y-) = \lim_{z \uparrow y} T^{\leftarrow}(z)$  and  $T^{\leftarrow}(y+) = \lim_{z \downarrow y} T^{\leftarrow}(z)$ .
- (GI7)  $T$  is continuous if and only if  $T^{\leftarrow} \uparrow$  on  $[\inf T, \sup T]$ .  
 $T \uparrow$  if and only if  $T^{\leftarrow}$  is continuous on  $\text{ran } T$ .
- (GI8) If  $T_1$  and  $T_2$  are right-continuous transformations with properties as  $T$ , then  $(T_1 \circ T_2)^{\leftarrow} = T_2^{\leftarrow} \circ T_1^{\leftarrow}$ .

*Proof.* See Embrechts and Hofert (2013). □

*Proof of Lemma 7.2.* Note that the *range of a rv*  $X$  is defined by

$$\text{ran } X = \{x \in \mathbb{R} : \mathbb{P}(X \in (x - h, x]) > 0 \text{ for all } h > 0\}.$$

Since  $F$  is continuous on  $\mathbb{R}$ , (GI7) implies that  $F^{\leftarrow} \uparrow$  on  $[\inf F, \sup F] = [0, 1]$ . Thus

$$\begin{aligned} \mathbb{P}(F(X) \leq u) &\stackrel{(GI7)}{=} \mathbb{P}(F^{\leftarrow}(F(X)) \leq F^{\leftarrow}(u)) \stackrel{(GI3)}{=} \mathbb{P}(X \leq F^{\leftarrow}(u)) \\ &= F(F^{\leftarrow}(u)) \stackrel{(GI4)}{=} u, \quad u \in [0, 1], \end{aligned}$$

where (GI3) applies since  $F \uparrow$  on  $\text{ran } X$ . □

*Proof of Lemma 7.6.*

$$\begin{aligned} \text{"}\Rightarrow\text{" } \mathbb{P}(F_j(X_j) \leq u_j \forall j) &\stackrel{\text{cont.}}{=} \mathbb{P}(F_j(X_j) < u_j \forall j) \stackrel{(GI5)}{=} \mathbb{P}(X_j < F_j^{\leftarrow}(u_j) \forall j) \\ &\stackrel{\text{cont.}}{=} \mathbb{P}(X_j \leq F_j^{\leftarrow}(u_j) \forall j) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)) \stackrel{\text{Sklar}}{=} C(\mathbf{u}). \end{aligned}$$

$\text{"}\Leftarrow\text{"}$  Since  $F_j \uparrow$  on  $\text{ran } X_j$ ,  $j \in \{1, \dots, d\}$ ,

$$\begin{aligned} F(\mathbf{x}) &\stackrel{(GI3)}{=} \mathbb{P}(F_j^{\leftarrow}(F_j(X_j)) \leq x_j \forall j) \stackrel{(GI5)}{=} \mathbb{P}(F_j(X_j) \leq F_j(x_j) \forall j) \\ &\stackrel{\text{ass.}}{=} C(F_1(x_1), \dots, F_d(x_d)) \stackrel{\text{Sklar}}{\Rightarrow} \mathbf{X} \text{ has copula } C \quad \square \end{aligned}$$

*Proof of Theorem 7.8.*

- 1) ■ By Lemma A.16,  $1 - C(\mathbf{u}) = C(\mathbf{1}) - C(\mathbf{u}) \leq \sum_{j=1}^d (1 - u_j) = d - \sum_{j=1}^d u_j$ , so  $C(\mathbf{u}) \geq \sum_{j=1}^d u_j - d + 1$ . Also,  $C(\mathbf{u}) \geq 0$ . So  $C(\mathbf{u}) \geq W(\mathbf{u})$ .
  - Since copulas are componentwise increasing,  $C(\mathbf{u}) \leq C(1, \dots, 1, u_j, 1, \dots, 1) = u_j$  for all  $j$ . Hence,  $C(\mathbf{u}) \leq \min_{1 \leq j \leq d} \{u_j\} = M(\mathbf{u})$ .
- 2)  $W$  is a copula for  $d = 2$  since  $(U, 1 - U) \sim W$  for  $U \sim U(0, 1)$ .  $W$  is not a copula for  $d \geq 3$  since

$$\begin{aligned}
 & \Delta_{(\frac{1}{2}, 1]} W \\
 &= \sum_{i \in \{0, 1\}^d} (-1)^{\sum_{j=1}^d i_j} W\left(\frac{1}{2}^{i_1}, \dots, \frac{1}{2}^{i_d}\right) \\
 &= \max\{1 + 1 + 1 + \dots + 1 - d + 1, 0\} \quad (i_j = 0 \ \forall j) \\
 & \quad - d \max\{\frac{1}{2} + 1 + 1 + \dots + 1 - d + 1, 0\} \quad (\exists! j : i_j = 1)
 \end{aligned}$$

$$\begin{aligned}
& + \binom{d}{2} \max\{\tfrac{1}{2} + \tfrac{1}{2} + 1 + \cdots + 1 - d + 1, 0\} \quad (\exists! \text{ two } j : i_j = 1) \\
& - \cdots + (-1)^d \max\{\tfrac{1}{2} + \cdots + \tfrac{1}{2} - d + 1, 0\} \quad (i_j = 1 \ \forall j) \\
& = 1 - \frac{d}{2} < 0 \quad \text{for } d \geq 3.
\end{aligned}$$

3)  $M$  is a copula for all  $d \geq 2$  since  $(U, \dots, U) \sim M$  for  $U \sim \text{U}(0, 1)$ .  $\square$

## Extreme value and Marshall–Olkin copulas

- *Extreme value copulas* are the *copulas*  $C$  of limiting distributions of properly location-scale transformed componentwise maxima of a sequence of random vectors.
- They are given by

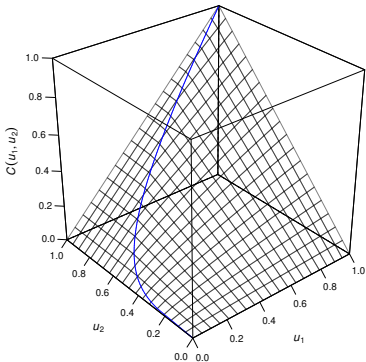
$$C(\mathbf{u}) = \left( \prod_{j=1}^d u_j \right)^{A\left(\frac{\log u_1}{\log \Pi(\mathbf{u})}, \dots, \frac{\log u_d}{\log \Pi(\mathbf{u})}\right)}$$

for a *Pickands dependence function*  $A$ ; see Ressel (2013) for a characterization of  $A$ .

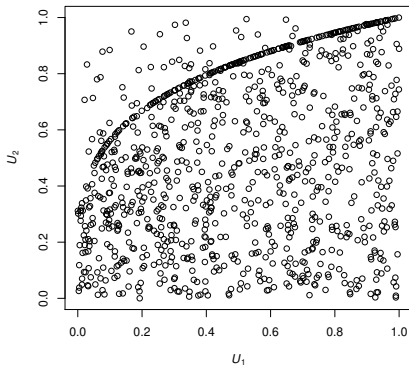
- **Examples:** Gumbel copula, Marshall-Olkin copulas.
- For more details, see Jaworski et al. (2010, Chapter 6).

Another class of copulas is given by  $C(u_1, u_2) = \min\{u_1 u_2^{1-\alpha_2}, u_1^{1-\alpha_1} u_2\}$ ,  $\alpha_1, \alpha_2 \in [0, 1]$ . Such copulas are called *Marshall–Olkin copulas* and one of their characteristics is a *singular component* (set of Lebesgue measure 0 in which  $(U_1, U_2)$  take values with a positive probability).

**MO copula with singular component** ( $\alpha_1 = 0.2, \alpha_2 = 0.8, \tau = 0.19$ )



**Scatter plot MO copula** ( $n = 1000, \alpha_1 = 0.2, \alpha_2 = 0.8, \tau = 0.19$ )





## Perfect dependence

*Proof of Proposition 7.14.* Consider Part 1); Part 2) works similarly.

“ $\Rightarrow$ ” By assumption,  $\mathbb{P}(X_2 \leq x)$  equals  $\mathbb{P}(F_2^{\leftarrow}(1 - F_1(X_1)) \leq x) \stackrel{(\text{GI5})}{=}$

$\mathbb{P}(1 - F_1(X_1) \leq F_2(x)) = F_2(x)$ . If  $(X_1, X_2)$  has copula  $C$ , then

$$C(\mathbf{u}) \stackrel{\text{L.7.6}}{\underset{\text{“only if”}}{=}} \mathbb{P}(F_1(X_1) \leq u_1, F_2(F_2^{\leftarrow}(1 - F_1(X_1))) \leq u_2)$$

$$\stackrel{(\text{GI4})}{=} \mathbb{P}(F_1(X_1) \leq u_1, 1 - F_1(X_1) \leq u_2)$$

$$= \mathbb{P}(1 - u_2 < U \leq u_1) = W(u_1, u_2) \quad \text{for } U \sim U(0, 1).$$

“ $\Leftarrow$ ”  $W(u_1, u_2) = 0$  for all  $u_1, u_2 \in [0, 1]$  such that  $u_1 + u_2 - 1 < 0$ , so  $W$  puts no mass below the secondary diagonal. Similarly one shows that  $W$  puts no mass above the diagonal. This implies that  $W$  puts mass 1 on the secondary diagonal. Since  $F_2 \uparrow \text{ran } X_2$ , we thus obtain  $\mathbb{P}(X_2 = F_2^{\leftarrow}(1 - F_1(X_1))) = \mathbb{P}(F_2(X_2) = F_2(F_2^{\leftarrow}(1 - F_1(X_1)))) \stackrel{(\text{GI4})}{=} \mathbb{P}(F_2(X_2) = 1 - F_1(X_1)) = \mathbb{P}(U_2 = 1 - U_1) = 1. \quad \square$

*Proof of Proposition 7.15.* Consider  $T(u) = F_1^{\leftarrow}(u) + \cdots + F_d^{\leftarrow}(u) \nearrow$ , left-continuous and let  $U \sim U(0, 1)$ . We first show that  $F_{T(U)}^{\leftarrow}(u) = T(u)$ , for all  $u \in [0, 1]$ .

$$1) \ T \text{ left-continuous} \Rightarrow T(u) \leq x \Leftrightarrow u \leq u_x := \sup\{u : T(u) \leq x\}$$

$$2) \ 1) \Rightarrow \{T(U) \leq x\} = \{U \leq u_x\} \Rightarrow F_{T(U)}(x) = F_U(u_x) = u_x.$$

$$\Rightarrow F_{T(U)}^{\leftarrow}(u) \leq x \stackrel{(GI5)}{\Leftrightarrow} F_{T(U)}(x) \geq u \stackrel{2)}{\Leftrightarrow} u_x \geq u \stackrel{1)}{\Leftrightarrow} T(u) \leq x$$

Choosing  $x = T(u)$  and  $x = F_{T(U)}^{\leftarrow}(u)$  in the last line, we see that  $F_{T(U)}^{\leftarrow}(u) = T(u)$ . Now Proposition 7.14 2) implies that

$$(X_1, \dots, X_d) \stackrel{d}{=} (F_1^{\leftarrow}(U), \dots, F_d^{\leftarrow}(U)),$$

so that

$$F_{\sum_{j=1}^d X_j}(x) = \mathbb{P}\left(\sum_{j=1}^d X_j \leq x\right) = \mathbb{P}\left(\sum_{j=1}^d F_j^{\leftarrow}(U) \leq x\right) = \mathbb{P}(T(U) \leq x)$$

$$\text{and thus } F_{\sum_{j=1}^d X_j}^{\leftarrow}(\alpha) = F_{T(U)}^{\leftarrow}(\alpha) = T(\alpha) = \sum_{j=1}^d F_j^{\leftarrow}(\alpha). \quad \square$$

## Linear correlation

*Proof of Proposition 7.16.* Let  $(X'_1, X'_2)$  be an iid-copy of  $(X_1, X_2)$ . Consider

$$\begin{aligned} & 2 \operatorname{cov}(X_1, X_2) \\ &= \mathbb{E}((X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)) + \mathbb{E}((X'_1 - \mathbb{E}X'_1)(X'_2 - \mathbb{E}X'_2)) \\ &= \mathbb{E}(((X_1 - \mathbb{E}X_1) - (X'_1 - \mathbb{E}X'_1)) \cdot ((X_2 - \mathbb{E}X_2) - (X'_2 - \mathbb{E}X'_2))) \\ &\stackrel{\text{check}}{=} \mathbb{E}((X_1 - X'_1)(X_2 - X'_2)). \end{aligned}$$

With  $b - a = \int_{-\infty}^{\infty} (I_{\{a \leq x\}} - I_{\{b \leq x\}}) dx$  for all  $a, b \in \mathbb{R}$ , we obtain that

$$\begin{aligned} & 2 \operatorname{cov}(X_1, X_2) \\ &= \mathbb{E} \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (I_{\{X'_1 \leq x_1\}} - I_{\{X_1 \leq x_1\}})(I_{\{X'_2 \leq x_2\}} - I_{\{X_2 \leq x_2\}}) dx_1 dx_2 \right] \\ &\stackrel{\text{Fubini}}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{E}(\dots) dx_1 dx_2 \stackrel{\text{multiply ind.}}{=} 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F(x_1, x_2) - F_1(x_1)F_2(x_2)) dx_1 dx_2. \end{aligned}$$

□

## Rank correlation

To overcome (some) of the deficiencies of  $\rho$ , Scarsini (1984) introduced:

### Definition A.18 (Rank correlation coefficient)

A measure of association  $\kappa = \kappa(X_1, X_2) = \kappa(C)$  between two continuously distributed random variables  $X_1$  and  $X_2$  with copula  $C$  is a *rank correlation coefficient (or measure of concordance)* if

- 1)  $\kappa$  exists for every pair  $(X_1, X_2)$  of cont. distributed random variables;
- 2)  $-1 \leq \kappa \leq 1$ ,  $\kappa(W) = -1$ , and  $\kappa(M) = 1$ ;
- 3)  $\kappa(X_1, X_2) = \kappa(X_2, X_1)$ ;
- 4)  $X_1$  and  $X_2$  being independent implies  $\kappa(X_1, X_2) = \kappa(\Pi) = 0$ ;
- 5)  $\kappa(-X_1, X_2) = -\kappa(X_1, X_2)$ ;
- 6)  $C_1(\mathbf{u}) \leq C_2(\mathbf{u})$  for all  $\mathbf{u} \in [0, 1]^2$  implies  $\kappa(C_1) \leq \kappa(C_2)$ ;
- 7)  $C_n \rightarrow C$  ( $n \rightarrow \infty$ ) pointwise implies  $\lim_{n \rightarrow \infty} \kappa(C_n) = \kappa(C)$ .

### Proposition A.19 (Basic properties of $\kappa$ )

Let  $\kappa$  be a rank correlation coefficient for two continuously distributed random variables  $X_1 \sim F_1$  and  $X_2 \sim F_2$ . Then

- 1)  $\kappa(X_1, X_2) = \kappa(C)$  ( $\kappa$  only depends on  $C$ ).
- 2) if  $T_j$  is a strictly increasing function on  $\text{ran } X_j$ ,  $j \in \{1, 2\}$ , then  $\kappa(T_1(X_1), T_2(X_2)) = \kappa(X_1, X_2)$ .

*Proof.*

- 1) Set  $(U_1, U_2) = (F_1(X_1), F_2(X_2))$ . By the invariance principle,  $(X_1, X_2)$  and  $(U_1, U_2)$  have the same copula  $C$ . Thus, by 6),  $\kappa(U_1, U_2) \leq \kappa(X_1, X_2)$ , but also  $\kappa(X_1, X_2) \leq \kappa(U_1, U_2)$ , so  $\kappa(X_1, X_2) = \kappa(U_1, U_2)$  ( $\Rightarrow$  only depends on  $C$ ).
- 2) Invariance principle  $\Rightarrow$  The copula  $C$  of  $(X_1, X_2)$  equals the copula of  $(T_1(X_1), T_2(X_2))$ . Hence  $\kappa(T_1(X_1), T_2(X_2)) = \kappa(C) = \kappa(X_1, X_2)$ .

□

## Kendall's tau and Spearman's rho

*Proof of Proposition 7.20.* Let  $(X'_1, X'_2)$  be an independent copy of  $(X_1, X_2)$ . Then

$$\begin{aligned}\rho_\tau &= \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) > 0) - \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) < 0) \\ &= 2 \underbrace{\mathbb{P}((X_1 - X'_1)(X_2 - X'_2) > 0)}_{=2\mathbb{P}(X_1 < X'_1, X_2 < X'_2)} - 1 = 4\mathbb{P}(U_1 \leq U'_1, U_2 \leq U'_2) - 1 \\ &= 4 \int_0^1 \int_0^1 \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2) dC(u_1, u_2) - 1\end{aligned}\quad \square$$

For computing  $\rho_\tau$ ,  $\int_{[0,1]^2} C(\mathbf{u}) d\tilde{C}(\mathbf{u}) = \frac{1}{2} - \int_{[0,1]^2} D_1 C(\mathbf{u}) D_2 \tilde{C}(\mathbf{u}) d\mathbf{u}$  is often helpful; see Li et al. (2002). One can also show that for any bivariate copulas  $C, \tilde{C}$ ,  $\int_{[0,1]^2} C(\mathbf{u}) d\tilde{C}(\mathbf{u}) = \int_{[0,1]^2} \tilde{C}(\mathbf{u}) dC(\mathbf{u})$ .

## Rank correlation for elliptical copulas

### Lemma A.20

Let  $\mathbf{X} \sim E_2(\mathbf{0}, \Sigma, \psi)$  with  $\mathbb{P}(\mathbf{X} = \mathbf{0}) = 0$  and  $\rho = P_{12} = \text{corr}(\Sigma)_{12}$ . Then

$$\mathbb{P}(X_1 > 0, X_2 > 0) = \frac{1}{4} + \frac{\arcsin \rho}{2\pi}.$$

*Proof.*

- Note that  $\mathbf{Y} = \begin{pmatrix} 1/\sqrt{\sigma_{11}} & 0 \\ 0 & 1/\sqrt{\sigma_{22}} \end{pmatrix} \mathbf{X} \sim E_2(\mathbf{0}, P, \psi)$  with  $P = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ .
- Let  $A = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}$  so that  $AA' = P$ . Then  $\mathbf{Y} \stackrel{d}{=} R\mathbf{A}\mathbf{U} \stackrel{d}{=} R\mathbf{A} \begin{pmatrix} \cos \Theta \\ \sin \Theta \end{pmatrix}$ ,  $\Theta \sim U(-\pi, \pi)$  independent of  $R$ .
- With  $\varphi = \arcsin \rho$ , we have  $\mathbf{Y} \stackrel{d}{=} R \begin{pmatrix} \cos \Theta \\ \sin \varphi \cos \Theta + \cos \varphi \sin \Theta \end{pmatrix} = \begin{pmatrix} \cos \Theta \\ \sin(\varphi + \Theta) \end{pmatrix}$ .
- Thus  $\mathbb{P}(X_1 > 0, X_2 > 0) = \mathbb{P}(Y_1 > 0, Y_2 > 0) = \mathbb{P}(\cos \Theta > 0, \sin(\varphi + \Theta) > 0) = \mathbb{P}(\Theta \in (-\frac{\pi}{2}, \frac{\pi}{2}), \varphi + \Theta \in (0, \pi)) = \mathbb{P}(\Theta \in (-\frac{\pi}{2}, \frac{\pi}{2}), \Theta \in (-\varphi, \pi - \varphi)) = \mathbb{P}(\Theta \in (-\varphi, \frac{\pi}{2}]) = (\frac{\pi}{2} - (-\varphi))/(2\pi). \quad \square$

### Lemma A.21 (Representation of Spearman's rho)

Let  $(U_1, U_2) \sim C$  and  $\tilde{U}_1, \bar{U}_2 \stackrel{\text{ind.}}{\sim} U(0, 1)$  be independent. Then  $\rho_S = \rho_S(U_1, U_2) = 12\mathbb{P}(U_1 \leq \tilde{U}_1, U_2 \leq \bar{U}_2) - 3$ .

*Proof.*  $12\mathbb{P}(U_1 \leq \tilde{U}_1, U_2 \leq \bar{U}_2) - 3 = 12\mathbb{E}(\mathbb{P}(\tilde{U}_1 > U_1, \bar{U}_2 > U_2 \mid U_1, U_2)) - 3 = 12\mathbb{E}((1 - U_1)(1 - U_2)) - 3 = 12\mathbb{E}(U_1 U_2) - 3 = \rho_S(U_1, U_2)$ .  $\square$   
ind.

*Proof of Proposition 7.26.*  $\mathbf{X} \stackrel{d}{=} \sqrt{W}\mathbf{Z}$  for  $\mathbf{Z} \sim N_2(\mathbf{0}, P)$ . Let  $\tilde{Z}, \bar{Z} \sim N(0, 1)$  and assume  $\mathbf{Z}, \tilde{Z}, \bar{Z}, W, \tilde{W}$  and  $\bar{W}$  are all independent. Let

$$\tilde{X} = \sqrt{\tilde{W}}\tilde{Z}, \quad \bar{X} = \sqrt{\bar{W}}\bar{Z},$$

$$Y_1 = X_1 - \tilde{X} = \sqrt{W}Z_1 - \sqrt{\tilde{W}}\tilde{Z},$$

$$Y_2 = X_2 - \bar{X} = \sqrt{W}Z_2 - \sqrt{\bar{W}}\bar{Z}.$$

$$\rho_S(X_1, X_2) \stackrel{\text{L. A.21}}{\underset{\Phi^{-1}}{=}} 12\mathbb{P}(X_1 \leq \tilde{X}_1, X_2 \leq \bar{X}_2) - 3$$



$$\begin{aligned}
&= 6\mathbb{P}((X_1 - \tilde{X}_1)(X_2 - \bar{X}_2) > 0) - 3 \\
&= 3(2\mathbb{E}(\mathbb{P}(Y_1 Y_2 > 0 \mid W, \tilde{W}, \bar{W})) - 1) \\
&= 3(4\mathbb{E}(\mathbb{P}(Y_1 > 0, Y_2 > 0 \mid W, \tilde{W}, \bar{W})) - 1).
\end{aligned}$$

Now note that  $\mathbf{Y} \mid W, \tilde{W}, \bar{W} \sim N_2(\mathbf{0}, \begin{pmatrix} W+\tilde{W} & W\rho \\ W\rho & W+\bar{W} \end{pmatrix})$  with  $\rho(Y_1, Y_2) = \frac{W\rho}{\sqrt{(W+\tilde{W})(W+\bar{W})}}$ . Apply Lemma A.20 to see that

$$\rho_S(X_1, X_2) = 3\left(4\mathbb{E}\left(\frac{1}{4} + \frac{\arcsin \rho}{2\pi}\right) - 1\right) = \frac{12}{2\pi}\mathbb{E}(\arcsin \rho(Y_1, Y_2)).$$

For Gauss copulas,  $F_W(x) = I_{\{x \geq 1\}}$ , thus  $W = \tilde{W} = \bar{W} = 1$  a.s. and the result follows.  $\square$

*Proof of Proposition 7.27.*

- Let  $(X'_1, X'_2)$  be an independent copy of  $(X_1, X_2)$ . We have already seen that  $\rho_\tau = 2\mathbb{P}((X_1 - X'_1)(X_2 - X'_2) > 0) - 1$ .
- With  $\mathbf{X} \stackrel{d}{=} R\mathbf{A}\mathbf{U}$  and  $\mathbf{X}' \stackrel{d}{=} R'\mathbf{A}\mathbf{U}' (\stackrel{d}{=} -\mathbf{X}')$  we have  $\mathbf{Y} = \mathbf{X} - \mathbf{X}' \stackrel{d}{=} \mathbf{0} + A(R\mathbf{U} - R'\mathbf{U}')$ . Note that the characteristic function of  $-\mathbf{X}'$  is

$\phi_{-X'}(\mathbf{t}) = \phi_{X'}(\mathbf{t}) = \phi_X(\mathbf{t})$  so that  $\phi_Y(\mathbf{t}) \stackrel{\text{ind.}}{=} \phi_X(\mathbf{t})\phi_{-X'}(\mathbf{t}) = \phi_X(\mathbf{t})^2$ ,  
hence  $\mathbf{Y} \sim E_2(\mathbf{0}, P, \psi^2)$ .

■ We thus obtain that

$$\begin{aligned}\rho_\tau &= 2\mathbb{P}(Y_1 Y_2 > 0) - 1 = 2(\mathbb{P}(Y_1 > 0, Y_2 > 0) + \mathbb{P}(Y_1 < 0, Y_2 < 0)) - 1 \\ &= 4\mathbb{P}(\mathbf{Y} > \mathbf{0}) - 1 \stackrel[\text{L.A.20}]{\text{cont.}}{=} \frac{2}{\pi} \arcsin \rho.\end{aligned}\quad \square$$

For a generalization to componentwise n.d.  $\mathbf{X}$ , see Lindskog et al. (2003).

## A.8 Aggregate risk

*Proof of Euler's rule.* Let  $f : \mathbb{R}^d \setminus \{\mathbf{0}\} \rightarrow \mathbb{R}$  be differentiable and positive homogeneous of order  $r > 0$ , so that

$$f(t\mathbf{x}) = t^r f(\mathbf{x}), \quad \text{for all } t > 0, \mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}.$$

Differentiating w.r.t.  $t$  (use the chain rule on the left-hand side) leads to  $rt^{r-1}f(\mathbf{x}) = \nabla f(t\mathbf{x})\mathbf{x} = \sum_{j=1}^d \frac{\partial f}{\partial x_j}(t\mathbf{x}) \cdot x_j$  and thus and thus (for  $t = 1$ )

$$rf(\mathbf{x}) = \nabla f(\mathbf{x})\mathbf{x} = \sum_{j=1}^d \frac{\partial f}{\partial x_j}(\mathbf{x}) \cdot x_j. \quad (136)$$

On the other hand, if  $f : \mathbb{R}^d \setminus \{\mathbf{0}\} \rightarrow \mathbb{R}$  fulfills (136) and is differentiable, consider  $g(t) = f(t\mathbf{x})$ ,  $t > 0$ . Then  $g$  satisfies

$$g'(t) = \nabla f(t\mathbf{x})\mathbf{x} = \frac{1}{t} \sum_{j=1}^d \frac{\partial f}{\partial x_j}(t\mathbf{x}) \cdot tx_j \stackrel{(136)}{=} \frac{1}{t} rf(t\mathbf{x}) = \frac{r}{t} g(t)$$

with the initial value  $g(1) = f(\mathbf{x})$ . The only solution is  $g(t) = t^r f(\mathbf{x})$ , so  $f(t\mathbf{x}) = t^r f(\mathbf{x})$  and thus  $f$  is homogeneous of order  $r > 0$ .  $\square$

# References

- Brockwell, P. J. and Davis, R. A. (1991), Time Series: Theory and Methods, 2nd, New York: Springer.
- Chavez-Demoulin, V., Embrechts, P., and Hofert, M. (2014), An extreme value approach for modeling operational risk losses depending on covariates, *Journal of Risk and Insurance*, to appear.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), Modelling Extremal Events for Insurance and Finance, Berlin: Springer.
- Embrechts, Paul and Hofert, Marius (2013), A note on generalized inverses, *Mathematical Methods of Operations Research*, 77(3), 423–432, doi: <http://dx.doi.org/10.1007/s00186-013-0436-7>.
- Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., eds. (2010), Copula Theory and Its Applications, vol. 198, Lecture Notes in Statistics – Proceedings, Springer.

- Kou, S. and Peng, X. (2014), On the Measurement of Economic Tail Risk, <http://arxiv.org/abs/1401.4787> (2014-06-09).
- Leadbetter, M. R. (1991), On a basis for Peaks over Threshold modeling, *Statistics and Probability Letters*, 12, 357–362.
- Li, X., Mikusiński, P., and Taylor, M. D. (2002), Some integration-by-parts formulas involving 2-copulas, *Distributions with Given Marginals and Statistical Modelling*, ed. by C. M. Cuadras, J. Fortiana, and J. A. Rodríguez-Lallena, Kluwer Academic Publishers, 153–159.
- Lindskog, F., McNeil, A. J., and Schmock, U. (2003), Kendall's tau for elliptical distributions, *Credit Risk: Measurement, Evaluation and Management*, ed. by G. Bol et al., Heidelberg: Physica-Verlag (Springer), 149–156.
- Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis*, London: Academic Press.
- Maronna, R. A. (1976), Robust M-Estimators of multivariate location and scatter, *The Annals of Statistics*, 4, 51–67.

- McNeil, A. J., Frey, R., and Embrechts, P. (2005), Quantitative Risk Management: Concepts, Techniques and Tools, Princeton: Princeton University Press.
- Ressel, P. (2013), Homogeneous distributions – And a spectral representation of classical mean values and stable tail dependence functions, *Journal of Multivariate Analysis*, 117, 246–256.
- Scarsini, M. (1984), On measures of concordance, *Stochastica*, 8(3), 201–218.