1 Risk Bounds with Known Marginal Distributions

As described in the introduction, a key problem of risk analysis is to derive (sharp) risk bounds on a portfolio $S = X_1 + \cdots + X_n$ under the given distributional information on a risk vector $X = (X_1, \ldots, X_n)$. In this chapter, we derive several explicit results for this problem under the assumption that only the marginal distributions F_j of X_j are known, but the dependence structure of X is completely unknown. In particular we introduce some basic notions of risk theory, such as worst case value-at-risk and tail value-at-risk portfolios, comonotonic risk vectors, the connection of upper risk bounds to convex ordering, and some basic results to obtain worst case value-at-risk portfolios. A more detailed presentation and extension of these results is given in Rüschendorf (2013, Chapters 2–4). Some detailed mixing results in Section 1.4 are due to several papers of Wang and coauthors (see Wang and Wang, 2011).

1.1 Some Basic Notions and Results of Risk Analysis: VaR, TVaR, Comonotonicity, and Convex Order

There are several risk measures of interest, like the value-at-risk (VaR), the tail value-at-risk (TVaR), and the classes of convex risk measures or of distortion risk measures. The VaR risk measure at level α, VaR_α, $\alpha \in (0,1)$ of the portfolio S is defined as the α-quantile of the distribution of S, i.e.,

$$\mathrm{VaR}_\alpha(S) = F_S^{-1}(\alpha) = \inf\{x \in \mathbb{R};\ \ F_S(x) \geq \alpha\}; \tag{1.1}$$

we also make use of the upper VaR as an upper α-quantile, i.e.,

$$\mathrm{VaR}_\alpha^+(S) = \sup\{x \in \mathbb{R};\ \ F_S(x) \leq \alpha\}. \tag{1.2}$$

The TVaR risk measure at level α, TVaR_α, takes into account also the magnitude of the risk above the α-quantile and is defined as

$$\mathrm{TVaR}_\alpha(S) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u^+(S)\,\mathrm{d}u = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(S)\,\mathrm{d}u. \tag{1.3}$$

From the definition it follows that TVaR is an upper bound of VaR, i.e., for $\alpha < 1$ it holds that

$$\mathrm{VaR}_\alpha(S) \leq \mathrm{VaR}_\alpha^+(S) \leq \mathrm{TVaR}_\alpha(S). \tag{1.4}$$

https://doi.org/10.1017/9781009367189.004 Published online by Cambridge University Press

In comparison to VaR, TVaR has the important property of being a convex risk measure. A risk measure ϱ is said to be **convex** (Föllmer and Schied, 2004) if it is monotone, translation invariant, and satisfies the important convexity condition,

$$\mathrm{TVaR}_\alpha(aX + (1-a)Y) \le a\,\mathrm{TVaR}_\alpha(X) + (1-a)\,\mathrm{TVaR}_\alpha(Y). \tag{1.5}$$

If the risk measure ϱ is also positive homogeneous, then it is called **coherent**.

Thus, using TVaR as a risk measure, a diversified portfolio is preferred concerning the magnitude of risk in comparison to an undiversified portfolio. The left TVaR measure at level α, LTVaR_α is similarly defined and considers the left tails (best case) of risks:

$$\mathrm{LTVaR}_\alpha(S) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_u(S)\,\mathrm{d}s. \tag{1.6}$$

An important property of a risk measure that is convex **law invariant**, i.e., one that only depends on the marginal distribution, is its consistency with respect to convex order $\le_{\mathrm{cx}}$.

Definition 1.1 (Convex order) Let X and Y be two random variables with finite means. X is smaller than Y in convex order, denoted by $X \le_{\mathrm{cx}} Y$, if for all convex functions f,

$$Ef(X) \le Ef(Y), \tag{1.7}$$

whenever both sides of (1.7) are well defined.

A law-invariant convex risk measure ϱ (e.g., TVaR) is consistent with respect to convex order on proper probability spaces such as L^1 (integrable random variables) and L^∞ (bounded random variables). In consequence it holds that $X \le_{\mathrm{cx}} Y$ implies

$$\mathrm{TVaR}_\alpha(X) \le \mathrm{TVaR}_\alpha(Y), \tag{1.8}$$

see Chapter 4 of Föllmer and Schied (2004), Jouini et al. (2006), Bäuerle and Müller (2006), and Burgert and Rüschendorf (2006). From this section on, we consider as the basic space of risks $\mathcal{X} = L^1$ and assume that all marginal distributions of a risk vector X have finite first moments when dealing with TVaR. For given distribution functions $F_1, \ldots, F_n$, let $\mathcal{F}(F_1, \ldots, F_n)$ denote the **Fréchet class** of all n-dimensional distribution functions F with marginal distribution functions $F_1, \ldots, F_n$. The classical Fréchet bounds characterize the Fréchet class $\mathcal{F}(F_1, \ldots, F_n)$.

Theorem 1.2 (Fréchet bounds)
a) For $F \in \mathcal{F}(F_1, \ldots, F_n)$ it holds that

$$\begin{aligned} F_-(x) &:= \Big(\sum_{i=1}^n F_i(x_i) - (n-1)\Big)_+ \le F(x) \\ &\le F_+(x) := \min_{1\le i\le n} F_i(x_i), \quad x \in \mathbb{R}^n. \end{aligned} \tag{1.9}$$

F_-, F_+ are called lower resp. upper Fréchet bounds (also called Hoeffding–Fréchet bounds).

b) $F_+ \in \mathcal{F}(F_1, \dots, F_n)$*; if* $n = 2$ *then* $F_- \in \mathcal{F}(F_1, F_2)$*.*

c) For a distribution function F *on* $\mathbb{R}^n$*, it holds that*

$$F \in \mathcal{F}(F_1, \dots, F_n) \iff F_- \leq F \leq F_+.$$

In particular, there exists for any n a largest distribution function with marginals F_i, the upper Fréchet bound F_+. For $n = 2$ there exists a smallest distribution function with marginals F_i, the lower Fréchet bound. In general, the upper and lower bounds in (1.9) are sharp. The upper bound F_+ is attained by the comonotonic risk vector.

Definition 1.3 (Comonotonicity, countermonotonicity)
Let $F_1, \dots, F_n$ be one-dimensional distribution functions, and let $U \sim U(0, 1)$ be uniformly distributed on $[0, 1]$. Then:

a)
$$X^c := (F_1^{-1}(U), \dots, F_n^{-1}(U)) \tag{1.10}$$

with $F_i^{-1}(\alpha) = \inf\{x \in \mathbb{R} \colon F_i(x) \geq \alpha\}$ is called a **comonotonic** risk vector.

b) For $n = 2$,

$$X_c := (F_1^{-1}(U), F_2^{-1}(1 - U)) \tag{1.11}$$

is called a **countermonotonic** (antimonotonic) risk vector.

Comonotonic risk vectors X are characterized by the fact that the components of X are ordered in the same way.

The co- resp. countermonotonic risk vectors realize the upper resp. lower Fréchet bounds F_+, F_-, i.e.,

$$X^c \sim F_+ \qquad \text{and for } n = 2, \quad X_c \sim F_-. \tag{1.12}$$

In terms of the lower orthant order $\leq_{\mathrm{lo}}$ defined by the pointwise ordering of the distribution functions, therefore, for any vector X with marginal distributions F_i it holds by the Fréchet bounds that

$$X \leq_{\mathrm{lo}} X^c \tag{1.13}$$

and for $n = 2$,

$$X_c \leq_{\mathrm{lo}} X. \tag{1.14}$$

The following basic result due to Meilijson and Nadas (1979) describes the role of the comonotonic vector as a worst case model for the portfolio $S = \sum_{i=1}^n X_i$ with respect to all law-invariant convex risk measures.

Theorem 1.4 (Comonotonic risk vector and convex order)
Let $X = (X_1, \dots, X_n)$ *be a risk vector with marginal distributions* F_i*. Then*

a)
$$\sum_{i=1}^n X_i \leq_{\mathrm{cx}} \sum_{i=1}^n X_i^c, \tag{1.15}$$

i.e., the portfolio of comonotonic risks is the worst case portfolio with respect to convex order.

https://doi.org/10.1017/9781009367189.004 Published online by Cambridge University Press

b)
$$E\Big(\sum_{i=1}^{n} X_i - t\Big)_+ \le E\Big(\sum_{i=1}^{n} X_i^c - t\Big)_+ \tag{1.16}$$

for all t. Moreover, $E\Big(\sum_{i=1}^{n} X_i^c - t\Big)_+ =: \Psi_+(t)$, where

$$\Psi_+(t) = \inf_{\sum v_i = t} \sum_{i=1}^{n} E(X_i - v_i)_+. \tag{1.17}$$

The statement in b) says that the excess of loss risk functional of the portfolio is maximized by the comonotonic risk vector.

For $n = 2$, a countermonotonic risk vector $X_c = (F_1^{-1}(U), F_2^{-1}(1-U))$ realizes the convex minimum of portfolio sums of variables X_i with distribution functions F_i.

Proposition 1.5 (Countermonotonic risk vector and convex order) *Let $X = (X_1, X_2)$ be a risk vector of size $n = 2$ with marginal distribution functions F_i. Then for all $X_i \sim F_i$ it holds that*

$$F_1^{-1}(U) + F_2^{-1}(1-U) \le_{\mathrm{cx}} X_1 + X_2. \tag{1.18}$$

In consequence, for $n = 2$ we have for all $X_i \sim F_i$,

$$X_{1,c} + X_{2,c} \le_{\mathrm{cx}} X_1 + X_2 \le_{\mathrm{cx}} X_1^c + X_2^c, \tag{1.19}$$

where $X_c = (X_{1,c}, X_{2,c})$.

We define the worst case risks of the portfolio $S = \sum_{i=1}^{n} X_i$, where X_i have marginal distribution functions F_i with respect to VaR and TVaR by

$$\begin{aligned} \overline{\mathrm{VaR}}_\alpha &:= \sup\Big\{\mathrm{VaR}_\alpha(S); S = \sum_{i=1}^{n} X_i, X_i \sim F_i, 1 \le i \le n\Big\} \\ \text{and}\quad \overline{\mathrm{TVaR}}_\alpha &:= \sup\Big\{\mathrm{TVaR}_\alpha(S); S = \sum_{i=1}^{n} X_i, X_i \sim F_i, 1 \le i \le n\Big\}. \end{aligned} \tag{1.20}$$

Similarly, the best case of risks at level α is defined as

$$\begin{aligned} \underline{\mathrm{VaR}}_\alpha &:= \inf\Big\{\mathrm{VaR}_\alpha(S); S = \sum_{i=1}^{n} X_i, X_i \sim F_i, 1 \le i \le n\Big\} \\ \text{and}\quad \underline{\mathrm{TVaR}}_\alpha &:= \inf\Big\{\mathrm{LTVaR}_\alpha(S); S = \sum_{i=1}^{n} X_i, X_i \sim F_i, 1 \le i \le n\Big\}. \end{aligned} \tag{1.21}$$

Then we get by means of Theorem 1.4 the following important connections between these notions. For a risk vector X, let $S = \sum_{i=1}^{n} X_i$ be the portfolio sum and $S^c = \sum_{i=1}^{n} X_i^c$ be the corresponding portfolio sum of the comonotonic risk vector X.

Theorem 1.6 *Let X be a risk vector with distribution function $F \in \mathcal{F}(F_1, \dots, F_n)$. Then for the portfolio $S = \sum_{i=1}^{n} X_i$, it holds that*

a)
$$\mathrm{VaR}_\alpha(S) \le \mathrm{TVaR}_\alpha(S) \le \mathrm{TVaR}_\alpha(S^c) = \sum_{i=1}^{n} \mathrm{TVaR}_\alpha(X_i), \tag{1.22}$$

b)
$$\sum_{i=1}^{n} \mathrm{LTVaR}_\alpha(X_i) = \mathrm{LTVaR}_\alpha(S^c) \le \mathrm{LTVaR}_\alpha(S) \le \mathrm{VaR}_\alpha(S), \tag{1.23}$$

https://doi.org/10.1017/9781009367189.004 Published online by Cambridge University Press

c)
$$\overline{\mathrm{VaR}}_\alpha \le \overline{\mathrm{TVaR}}_\alpha = \sum_{i=1}^n \mathrm{TVaR}_\alpha(X_i)$$
$$\text{and} \quad \underline{\mathrm{LTVaR}}_\alpha = \sum_{i=1}^n \mathrm{LTVaR}_\alpha(X_i) \le \underline{\mathrm{VaR}}_\alpha, \tag{1.24}$$

d)
$$\mathrm{VaR}_\alpha(S^c) = \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i). \tag{1.25}$$

Proof The inequality $\mathrm{VaR}_\alpha(S) \le \mathrm{TVaR}_\alpha(S)$ is immediate from the definition of $\mathrm{TVaR}_\alpha(S)$. Since TVaR_α is a convex law-invariant risk measure, we obtain the inequality $\mathrm{TVaR}_\alpha(S) \le \mathrm{TVaR}_\alpha(S^c)$ by the consistency with respect to convex order from Theorem 1.4.

Using that

$$\alpha\, \mathrm{LTVaR}_\alpha(S) + (1-\alpha)\, \mathrm{TVaR}_\alpha(S) = ES, \tag{1.26}$$

we obtain

$$\mathrm{LTVaR}_\alpha(S^c) = \sum_{i=1}^n \mathrm{LTVaR}_\alpha(X_i) \le \mathrm{LTVaR}_\alpha(S) \le \mathrm{VaR}_\alpha(S).$$

Finally for $S^c = \sum_{i=1}^n F_i^{-1}(U)$, it holds by the comonotonicity of the summands:

$$S^c \ge \mathrm{VaR}_\alpha(S^c)$$

if and only if for all i, $X_i = F_i^{-1}(U) \ge \mathrm{VaR}_\alpha(X_i)$, i.e.,

$$\mathrm{VaR}_\alpha(S^c) = \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i), \tag{1.27}$$

$$\text{and} \quad \mathrm{TVaR}_\alpha(S^c) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(S^c)\,\mathrm{d}u \tag{1.28}$$
$$= \frac{1}{1-\alpha}\int_\alpha^1 \sum_{i=1}^n \mathrm{VaR}_u(X_i)\,\mathrm{d}u = \sum_{i=1}^n \mathrm{TVaR}_\alpha(X_i). \qquad \square$$

Remark 1.7 The inequalities (1.22) and (1.24) give a simple way to calculate an upper bound for the worst case VaR, whereas inequality (1.23) gives a lower bound for the best case VaR. The VaR of the comonotonic risk portfolio is easy to calculate, but it turns out that it is not a worst case with respect to VaR. The comonotonic portfolio is, however, a worst case portfolio with respect to TVaR, and hence the worst case TVaR bound is easy to determine. ◊

1.2 Standard Bounds, VaR Bounds, and Worst Case Distributions

It is an important task to describe good upper bounds for the value-at-risk and to determine worst case portfolios. The insight that the comonotonic portfolio is not the worst case VaR portfolio was a surprise in the practice of risk analysis and led to a rethinking of basic recommendations in risk regulation.

The standard bounds for the distribution function of the sum

$$M_n^{\le}(t) = \sup\Big\{P\Big(\sum_{i=1}^n X_i \le t\Big); X_i \sim F_i,\ \ 1 \le i \le n\Big\},$$
$$m_n^{\le}(t) = \inf\Big\{P\Big(\sum_{i=1}^n X_i \le t\Big); X_i \sim F_i,\ \ 1 \le i \le n\Big\},$$

resp. for the corresponding tail risks

$$M_n(t) = \sup\Big\{P\Big(\sum_{i=1}^n X_i \ge t\Big); X_i \sim F_i,\ \ 1 \le i \le n\Big\},$$
$$m_n(t) = \inf\Big\{P\Big(\sum_{i=1}^n X_i \ge t\Big); X_i \sim F_i,\ \ 1 \le i \le n\Big\},$$

have been known in the literature for a long time, see Sklar (1973), Moynihan et al. (1978), Denuit et al. (1999), and Rüschendorf (2005).

Theorem 1.8 (Standard bounds) *Let $X = (X_1, \dots, X_n)$ be a random vector with marginal distribution functions $F_1, \dots, F_n$. Then for any $t \in \mathbb{R}$, it holds that*

$$\begin{aligned}\Big(\bigvee_{i=1}^n F_i(t) - (n-1)\Big)_+ &\le P\Big(\sum_{i=1}^n X_i \le t\Big)\\ &\le \min\Big(\bigwedge_{i=1}^n F_i(t), 1\Big),\end{aligned} \tag{1.29}$$

where $\bigwedge_{i=1}^n F_i(t) = \inf\{\sum_{i=1}^n F_i(u_i); \sum_{i=1}^n u_i = t\}$ is the "infimal convolution" of the (F_i), and $\bigvee_{i=1}^n F_i(t) = \sup\{\sum_{i=1}^n F_i(u_i); \sum_{i=1}^n u_i = t\}$ is the "supremal convolution" of the (F_i).

Proof For any $u_1, \dots, u_n$ with $\sum_{i=1}^n u_i = t$, it holds that

$$\begin{aligned}P\Big(\sum_{i=1}^n X_i \le t\Big) &\ge P\Big(\bigcup_{i=1}^n (X_i \le u_i)\Big),\\ &\ge \sum_{i=1}^n F_i(u_i),\end{aligned} \tag{1.30}$$

which implies the upper bound in (1.29). Similarly, using the Fréchet lower bound in (1.9) we obtain

$$\begin{aligned}P\Big(\sum_{i=1}^n X_i \le t\Big) &\ge P\Big(X_1 \le u_1, \dots, X_n \le u_n\Big)\\ &\ge \Big(\sum_{i=1}^n F_i(u_i) - (n-1)\Big)_+.\end{aligned} \tag{1.31}$$

□

https://doi.org/10.1017/9781009367189.004 Published online by Cambridge University Press

In general, the standard bounds in Theorem 1.8 are not sharp and can be considerably improved. Define for general n,

$$A_n(t) := \Big\{(x_1, \dots, x_n); \sum_{i=1}^{n} x_i \le t\Big\},$$

$$A_n^+(t) := \Big\{(x_1, \dots, x_n); \sum_{i=1}^{n} x_i < t\Big\}, \quad t \in \mathbb{R}^1,$$

and let

$$(F_1 \wedge F_2)^-(t) = \inf\{F_1(x-) + F_2(t-x); \quad x \in \mathbb{R}^1\}$$

denote the left continuous version of $F_1 \wedge F_2$; similarly, let $(F_1 \vee F_2)^-(t)$ be the left continuous version of $F_1 \vee F_2$. In the case $n = 2$, it was proved independently in Makarov (1981) and Rüschendorf (1982) that the standard bounds are sharp.

Theorem 1.9 (Sharpness of standard bounds, $n = 2$) *If X_i have distribution functions F_i, $1 = 1, 2$, then*

$$P(X_1 + X_2 \le t) \le M_2^{\le}(t) = (F_1 \wedge F_2)^-(t) \tag{1.32}$$

and

$$P(X_1 + X_2 < t) \ge m_2^{<}(t) = ((F_1 \vee F_2)^-(t) - 1)_+. \tag{1.33}$$

The proof of Theorem 1.9 given in Makarov (1981) uses direct arguments on the copulas, while the proof in Rüschendorf (1982) is based on duality theory. This latter proof allows us also to determine the worst case dependence structure.

On the unit interval [0, 1] supplied with the Lebesgue measure λ, define the random variables

$$Y_1(s) = F_1^{-1}(s), \quad Y_2(s) = F_2^{-1}(\varphi(S)), \tag{1.34}$$

with $\varphi(s) = 1 - s$, $0 \le s \le h(t)$, and $\varphi(s) = s$, $h(t) \le s \le 1$. Then the random variables Y_1, Y_2 maximize the distribution function of the sum at point t, i.e., they maximize $P(X_1 + X_2 < t)$. This means that they minimize the tail risk $P(X_1 + X_2 \ge t)$.

Proposition 1.10 (Maximizing (best case) pairs) *The random variables defined in* (1.34) *satisfy:*

a) $$Y_1 \sim F_1, \quad Y_2 \sim F_2,$$

b) $$P(Y_1 + Y_2 \le t) = M_2^{\le}(t) = (F_1 \wedge F_2)^-(t). \tag{1.35}$$

Proof The Lebesgue measure λ is invariant with respect to φ, i.e., $\lambda^\varphi = \lambda$. Therefore, $\lambda^{Y_i} = \lambda^{f_i^{-1} \circ \varphi} = \lambda^{F_i^{-1}}$, and thus $Y_i \sim F_i$, $i = 1, 2$. Since $F_i^{-1} \circ F_i(x) \le x$, we obtain for $s = F_1(u)$,

$$\begin{aligned} F_1^{-1}(s) + F_2^{-1}(h(t) - s) &= F_1^{-1} \circ F_1(u) + F_2^{-1}(h(t) - F_1(u)) \\ &= u + F_2^{-1}(F_2(t-u)) \le u + (t-u) = t. \end{aligned}$$

https://doi.org/10.1017/9781009367189.004 Published online by Cambridge University Press

For the sup in the definition of $g(t) = F_1^{-1} \vee F_2^{-1}(t)$, it is enough to consider s of the form $F_1(u)$. This implies $g \circ h(t) \leq t$, and it follows that

$$\lambda(\{Y_1 + Y_2 \leq t\}) \geq h(t) = (F_1 \wedge F_2)^-(t).$$

This implies b) and moreover

$$h(t) = \lambda(\{Y_1 + Y_2 \leq t\}) = F_1^{-1} \vee F_2^{-1}(t) = g(t). \qquad \square$$

A similar construction yields a worst case pair of risks minimizing the probability $P(X_1 + X_2 < t)$ or equivalently maximizing $P(X_1 + X_2 \geq t)$.

Define $Y_1(s) = F_1^{-1}(s)$, $Y_2(s) = F_2^{-1}(\varphi(s))$, $s \in [0,1]$ with

$$\varphi(s) = s, \quad 0 \leq s \leq h(t) \quad \text{and} \quad \varphi(s) = 1 - s, \quad h(t) \leq s \leq 1. \tag{1.36}$$

Then Y_1, Y_2 are obtained by a countermonotonic coupling in the upper part of the distributions, and we obtain in a similar way as in Proposition 1.10 that the risk variables Y_1, Y_2 determine a worst case pair of risks.

Proposition 1.11 (Worst case risks) *The random variables Y_1, Y_2 defined in* (1.36) *satisfy*

a) $Y_1 \sim F_1, \quad Y_2 \sim F_2$

and

b)
$$\begin{aligned} P(Y_1 + Y_2 \geq t) = M_2(t) &= 1 - m_2^<(t) \\ &= \min(2 - (F_1 \vee F_2)^-(t), 1) \\ &= \sup\{P(X_1 + X_2 \geq t); X_i \sim F_i\}. \end{aligned} \tag{1.37}$$

Definition 1.12 (Rearrangements) Let f, g be measurable, real functions on $[0,1]$. Then f is a rearrangement of g (with respect to the Lebesgue measure λ), notation $f \sim_r g$, if $\lambda^f = \lambda^g$, i.e., f, g have the same distribution with respect to λ.

The best resp. worst case couplings with respect to the tail risk can also be described by rearrangements.

Corollary 1.13 (Best and worst case risks by rearrangements) *For any $t \in \mathbb{R}$, it holds that*

a)
$$\begin{aligned} \alpha^* = M_2^{\leq}(t) &= \sup\{\alpha \in [0,1] \colon \exists f_j^\alpha \sim_r F_j^{-1} \textit{ on } [0,\alpha], \\ &\qquad \textit{such that } f_1^\alpha(s) + f_2^\alpha(s) \leq t, \textit{ for all } s \in [0,\alpha]\} \\ &= \inf\{\alpha \in [0,1] \colon \exists f_j^\alpha \sim_r F_j^{-1} \textit{ on } [\alpha,1], \\ &\qquad \textit{such that } f_1^\alpha(s) + f_2^\alpha(s) > t, \textit{ for all } s \in [\alpha,1]\} \end{aligned} \tag{1.38}$$

b)
$$\begin{aligned} \beta^* = M_2(t) &= \sup\{P(X_1 + X_2 \geq t), X_i \sim F_i, i = 1,2\} \\ &= \inf\{\alpha \in [0,1] \colon \exists f_j^{-1} \sim_r F_j^{-1} \textit{ on } [\alpha,1], \\ &\qquad \textit{such that } f_1^\alpha(s) + f_2^\alpha(s) \geq t, \textit{ for all } s \in [\alpha,1]\}. \end{aligned} \tag{1.39}$$

This rearrangement description also characterizes worst and best case couplings in the general case $n \geq 2$ (see Rüschendorf, 1983a; Puccetti and Rüschendorf, 2012a).

https://doi.org/10.1017/9781009367189.004 Published online by Cambridge University Press

Theorem 1.14 (Structure of worst and best case couplings) *For all distribution functions $F_1, \dots, F_n$ and $t \in \mathbb{R}$, it holds that*

$$M_n^{\le}(t) = \sup \Big\{\alpha \in [0,1];\ \textit{there exist } f_j^\alpha \sim_r F_j^{-1} \textit{ on } [0,\alpha], \quad 0 \le j \le n,\ \textit{such that } \sum_{j=1}^n f_j^\alpha \le t \textit{ on } [0,\alpha]\Big\}. \tag{1.40}$$

Similarly,

$$M_n(t) = 1 - m_n(t) = \inf \Big\{\alpha \in [0,1];\ \textit{there exist } f_j^\alpha \sim F_j^{-1} \textit{ on } [\alpha,1], \quad \textit{such that } \sum_{j=1}^n f_j^\alpha \ge t \textit{ on } [\alpha,1]\Big\}. \tag{1.41}$$

By Theorem 1.14, the problem of getting sharp bounds on the distribution function of the sum is reduced to a rearrangement problem. This rearrangement formulation motivates the construction of a fast algorithm to approximate the sharp bounds numerically – the rearrangement algorithm (RA) (see Puccetti and Rüschendorf, 2012b and Chapter 3). The proposed algorithm works well for general inhomogeneous portfolios, also those with high dimensions (i.e., d in the thousands).

The results of Propositions 1.10 and 1.11 imply that for $n = 2$, the worst case distribution maximizing $M_n(t)$ resp. maximizing $\mathrm{VaR}_\alpha(X_1 + X_2)$ is obtained by the countermonotonic coupling in the corresponding upper part of the distributions (see (1.36) and (1.37)), and the best case distribution minimizing $M_n(t)$ resp. minimizing $\mathrm{VaR}_\alpha(X_1 + X_2)$ is obtained by the countermonotonic coupling in the lower part of the distributions (see (1.34)).

Let $F_i^\alpha (F_{i,\alpha})$ denote the distribution F_i restricted to the upper (lower) α-part of F_i, i.e., formally $F_i^\alpha (F_{i,\alpha})$ is the distribution of $F_i^{-1}(U)$, where U is uniformly distributed on $[\alpha, 1]$ ($[0,\alpha]$). Then by Propositions (1.4) and (1.5), the worst case distribution (resp. best case distribution) minimizes (resp. maximizes) the distribution function of the portfolio sum in the upper part (resp. in the lower part) of the distribution. In other words, the upper resp. the lower parts of the distributions are flattened as much as possible. This principle also extends to $n \ge 2$: see Bernard et al. (2017c, Theorem 2.5).

Theorem 1.15 (VaR-bounds and convex order) *Let F_i^α denote the upper α-part of F_i. Then*

a)
$$\overline{\mathrm{VaR}}_\alpha^+ = \sup_{X_i \sim F_i} \mathrm{VaR}_\alpha^+\Big(\sum_{i=1}^n X_i\Big) = \sup_{Y_i^\alpha \sim F_i^\alpha} \mathrm{VaR}_0^+\Big(\sum_{i=1}^n Y_i^\alpha\Big). \tag{1.42}$$

b) If $X_i^\alpha, Y_i^\alpha \sim F_i^\alpha$ and

$$S^\alpha = \sum_{i=1}^n Y_i^\alpha \le_{\mathrm{cx}} \sum_{i=1}^n X_i^\alpha, \textit{ then} \tag{1.43}$$

$$\mathrm{VaR}_0^+\Big(\sum_{i=1}^n X_i^\alpha\Big) \le \mathrm{VaR}_0^+\Big(\sum_{i=1}^n Y_i^\alpha\Big).$$

https://doi.org/10.1017/9781009367189.004 Published online by Cambridge University Press

If it is possible to minimize the sum in the upper part of the distributions, i.e., for F_i^α in convex order, then one obtains as in the case $n = 2$ a worst case joint distribution, maximizing the VaR. This minimization of the convex order in the upper part is achieved in particular in the mixing case.

Definition 1.16 (Mixability)

a) Distribution functions $F_1, \dots, F_n$ on $\mathbb{R}$ are called **mixable** if there exist $X_i \sim F_i$ such that $\sum_{i=1}^n X_i = \mu$ a.s. for some $\mu \in \mathbb{R}^1$.
b) A distribution function F on $\mathbb{R}$ is called ***n*-mixable** (with center μ) if $F_1 = F, \dots, F_n = F$ are mixable.

Since for n-mixable distributions the mixing variables realize the convex minimum, we obtain as an immediate consequence of Theorem 1.15:

Corollary 1.17 *If $Y_i^\alpha \sim F_i^\alpha$, $1 \le i \le n$ exist, such that $S^\alpha = \sum_{i=1}^n Y_i^\alpha = c$, i.e., (F_i^α), $1 \le i \le n$ are mixable, then for all $X_i \sim F_i$, $1 \le i \le n$, it holds that*

$$\mathrm{VaR}_\alpha^+\Big(\sum_{i=1}^n X_i\Big) \le \mathrm{VaR}_0^+(S^\alpha) = c. \tag{1.44}$$

Remark 1.18

a) As stated in (1.43), the worst value for $\mathrm{VaR}_\alpha(S)$ is attained by the lower support point of some minimal element in convex order in this class $\mathcal{F}^\alpha = \{\sum_{i=1}^n Y_i; Y_i \sim F_i^\alpha\}$. For $d = 2$, a smallest element in this class exists and is given by the countermonotonic pair $Y_1^\alpha = (F_1^\alpha)^{-1}(U)$, $Y_2^\alpha = (F_2^\alpha)^{-1}(1-U)$, where U is uniformly distributed on $[\alpha, 1]$. The resulting $\mathrm{VaR}_0^+(Y_1^\alpha + Y_2^\alpha)$ is by Proposition 1.11 a sharp upper bound and is identical to the solution of this case in Rüschendorf (1982).
b) Similarly to Theorem 1.15, we obtain a corresponding result for the lower bound for $\mathrm{VaR}_\alpha\left(\sum_{i=1}^n X_i\right)$. Let $F_{i,\alpha}$ denote the distributions F_i restricted to the lower α-part.
For $X_{i,\alpha}, Y_{i,\alpha} \sim F_{i,\alpha}$ we get: If $S_\alpha = \sum_{i=1}^n Y_{i,\alpha} \le_{\mathrm{cx}} \sum_{i=1}^n X_{i,\alpha} = S'_\alpha$, then

$$\mathrm{VaR}_1(S_\alpha) \le \mathrm{VaR}_1(S'_\alpha). \tag{1.45}$$

In consequence we obtain: If S_α is a smallest sum with respect to $(F_{i,\alpha})$ in convex order, then

$$\mathrm{VaR}_1(S_\alpha) = \inf\Big\{\mathrm{VaR}_\alpha\Big(\sum_{i=1}^n X_i\Big); X_i \sim F_i\Big\}. \tag{1.46}$$

c) n-mixability has been established for uniform $U(0,1)$-distributions and binomial distributions in Gaffke and Rüschendorf (1981) and Rüschendorf (1982, 1983b), and for symmetric unimodal distributions in Rüschendorf and Uckelmann (2002). Wang and Wang (2011) established n-mixability for distributions with a decreasing density on $[0,1]$ under a moderate moment condition. Mixing in the case of concave densities on an interval (a,b) was established in Puccetti et al. (2012). For n small, say $n = 2, 3$, mixing is a rare property while for large enough n, by the abovementioned results mixing typically holds on bounded domains under monotonicity or concavity conditions. For more details on mixing, see Section 1.4. ◊

1.3 Worst Case Risk Vectors: The Conditional Moment Method and the Mixing Method

The conditional moment method gives an upper bound on the tail risk of the portfolio in terms of conditional moments of the marginals. Combined with a mixing condition on the marginal distributions, the upper bound is attained. This method was introduced in the case of homogeneous portfolios with monotone densities on $[0, 1]$ in Wang and Wang (2011). It was extended to general inhomogeneous distributions $F_1, \dots, F_n$ in Puccetti and Rüschendorf (2012b) and in Wang et al. (2013). In fact, it turns out that the conditional moment bounds have to be improved to be attainable. This improvement is given in a direct constructive way in Wang et al. (2013), while it is obtained in Puccetti and Rüschendorf (2012b) based on duality theory (see Chapter 3).

Let $X_i \sim F_i$, $G_i = F_i^{-1}$, the generalized inverse of F_i, $G = \sum_{i=1}^n G_i$, and assume that $\mu_i = EX_i$ exists. For $a \in [0, 1]$ define $\Psi(a)$ as the sum of the conditional first moments, given $X_i \geq G_i(a)$, i.e.,

$$\Psi(a) = \frac{1}{1-a} \int_a^1 G(t)\,\mathrm{d}t = \sum_{i=1}^n E(X_i \mid X_i \geq G_i(a)). \tag{1.47}$$

Then Ψ is monotonically non-decreasing and $\Psi(0) = \mu = \sum_{i=1}^n \mu_i$.

Theorem 1.19 (Method of conditional moments) *Let $X_i \sim F_i$ have first moments μ_i, $1 \leq i \leq n$. Then, for $s \geq \mu$, we have*

$$M_n(s) = \sup\Big\{P\Big(\sum_{i=1}^n X_i \geq s\Big);\ \ X_i \sim F_i, 1 \leq i \leq n\Big\} \leq 1 - \Psi^-(s), \tag{1.48}$$

where $\Psi^-(s) = \sup\{t \in [0, 1]; \Psi(t) \leq s\}$ is the left-continuous generalized inverse of Ψ.

Proof With $X_i \sim F_i$ and $S = \sum_{i=1}^n X_i$, we have

$$\begin{aligned}\mu = \sum_{i=1}^n \mu_i &= E[S] \geq E\left[S 1_{\{S<s\}}\right] + sP(S \geq s) \qquad (1.49)\\ &= \int_0^{P(S<s)} G(t)\,\mathrm{d}t + sP(S \geq s) = \mu - \int_{P(S<s)}^1 G(t)\,\mathrm{d}t + sP(S \geq s).\end{aligned}$$

If $P(S \geq s) > 0$, this implies that $\Psi(P(S < s)) \geq s$ and thus

$$P(S < s) \geq \Psi^-(s).$$

As a consequence, we obtain

$$P(S \geq s) \leq 1 - \Psi^-(s). \qquad \square$$

Remark 1.20

a) **Sharpness of conditional moment bounds.** The conditional bound in (1.48) is sharp if and only if the estimate in (1.49) is an equality, that is, if for the optimal

https://doi.org/10.1017/9781009367189.004 Published online by Cambridge University Press

coupling it holds true that $\{S \geq s\} = \{S = s\}$ a.s. This means, by Theorem 1.14, that the corresponding optimal rearrangements f_i^α on $[\alpha, 1]$ satisfy

$$\sum_{i=1}^{n} f_i^\alpha(u) = s \quad \text{for all} \quad u \in [\alpha, 1],$$

with $1 - \alpha = M(s)$, i.e., the random variables are mixing on the upper part of the distribution.

b) For unbounded domains, the bound in (1.48) typically fails to be sharp. To be a good bound it is indeed necessary that

$$\sum_{i=1}^{n} E\left(X_i \mid X_i \geq G_i(\alpha)\right) \approx s.$$

c) The method to get upper bounds for $M(s)$ implies directly also a lower bound for $P(S > s)$. Denoting by H the conditional moment function associated with the random variable $-X_i$, we obtain

$$P(S > s) = 1 - P((-S) \geq (-s)) \geq H^{-1}(-s).$$

In fact the conditional moment bound in Theorem 1.19 for the tail risk is equivalent to the TVaR bound for VaR in (1.22). The sharpness statement in Remark 1.20 a) corresponds to the sharpness of the bounds in Corollary 1.17,

$$\mathrm{VaR}_\alpha^+\Big(\sum_{i=1}^{n} X_i\Big) \leq \mathrm{VaR}_0^+(S^\alpha) = c, \tag{1.50}$$

under the mixing condition on the $\{F_i^\alpha\}$, $1 \leq i \leq n$. Note that under this condition $\mathrm{VaR}_0^+(S^\alpha) = c = \mathrm{TVaR}_\alpha(S)$. ◊

A modification of the method of conditional moments in Theorem 1.19 as in Remark 1.20 allows us to give improved bounds and even sharpness results not only for bounded domains but also for unbounded domains. Define for $s \geq \mu$ and $t \in [0, 1]$ the function $H_t(t_1)$ as the conditional expected moment function on the interval $[t, t_1]$, i.e.,

$$H_t(t_1) = \frac{1}{t_1 - t}\int_t^{t_1} G(u)\,\mathrm{d}u = EG(U_{[t,t_1]}). \tag{1.51}$$

Here $U_{[t,t_1]}$ denotes a random variable uniformly distributed on $[t, t_1]$. H_t is increasing in t, t_1. Let $H_t(1) \geq s$ and $G(t) \leq s$. This allows us to define

$$t_1 = t_1(t) = H_t^{-1}(s). \tag{1.52}$$

If we assume continuity of the F_i, then we get that the conditional expectation on $[t, t_1(t)]$ is identical to s:

$$H_t(t_1(t)) = s. \tag{1.53}$$

Without continuity we postulate (1.53) for the risk level considered. Next we define the optimal choice of such t's with (1.52) and (1.53):

$$t_0 = t_0(s) = \inf\{t;\ \ G_i \mid [t, t_1(t)], 1 \leq i \leq n \text{ are mixing with value } s\}, \tag{1.54}$$

https://doi.org/10.1017/9781009367189.004 Published online by Cambridge University Press

that is, t_0 is the infimum of all those t's such that there exist rearrangements $f_i^t \sim_r G_i \mid [t, t_1(t)]$, which satisfy

$$\sum_{i=1}^n f_i^t = E \sum_{i=1}^n G_i \mid [t, t_1(t)] = s. \tag{1.55}$$

Proposition 1.21 (Extended conditional moment method) *Let $X_i \sim F_i$, $1 \le i \le n$, be risk variables, and assume that the mixing condition* (1.55) *holds for some t. Then for $s \ge \mu$ we obtain the upper tail risk bound*

$$M_n(s) = \sup \Big\{P\Big(\sum_{i=1}^n X_i \ge s\Big);\ \ X_i \sim F_i\Big\} \le 1 - t_0(s). \tag{1.56}$$

Proof Under the "mixing assumption" (1.55) to $t_0 \in [0,1]$, we have $t_1(t_0) > t_0$, and the restricted distributions of $(G_i \mid [t_0, t_1(t_0)])$ are mixing. Therefore, there exist $\widetilde{V}_i \sim U_{[t_0, t_1(t_0)]}$ such that $\sum_{i=1}^n G_i(\widetilde{V}_i) = s$. Consequently this implies the existence of $V_i \sim U_{[t_0,1]}$ such that

$$\sum_{i=1}^n G_i(V_i) \ge s, \tag{1.57}$$

and the result follows. □

In a second step, the bound in (1.56) is further improved. This improvement has been shown in general classes of examples to yield sharp VaR bounds in Puccetti and Rüschendorf (2012a), Wang et al. (2013, Section 2.3), and Wang (2015).

Theorem 1.22 (Improved extended conditional moment bounds) *Assume that the risk variables $X_i \sim F_i$, $1 \le 1 \le n$ satisfy the mixing condition* (1.55)*, with $t_0 < 1$ and let $t_1 = t_1(t_0)$. Define*

$$t_2 = t_2(s) = \inf \Big\{t \le t_0;\ \textit{there exists a coupling}\ U_i \sim U_{[t,t_0]\cup[t_1,1]}\ \textit{with}\ \sum_{i=1}^n G_i(U_i) \ge s\Big\}. \tag{1.58}$$

Then

$$M_n(s) \le 1 - t_2(s). \tag{1.59}$$

Proof The admissible coupling $(\widetilde{V}_i)$ on $[t_0, t_1]$ in the proof of Proposition 1.21 can be improved using the admissible coupling (U_i) on $[t_2, t_0] \cup [t_1, 1]$ to an admissible coupling, say (V_i) on $[t_2, 1]$, satisfying $\sum_{i=1}^n G_i(V_i) \ge s$ on $[t_2, 1]$; this implies (1.59). □

Remark 1.23 (Structure of the worst case risks) The structure of the "optimal" coupling yielding worst case portfolios thus has a mixing part on $[t_0, t_1]$ and is admissible on $[t_2, 1]$. In the homogeneous case $F_i = F$, $1 \le i \le n$, it turns out that for the admissible part on $[t_2, t_0] \cup [t_1, 1]$ it is often sufficient to couple one "large" observation corresponding to $u \in [t_1, 1]$ with $n-1$ "small" observations in $[t_2, t_0]$ chosen to be identical (see the following discussion of this structure and the approach via duality in Chapter 3). ◊

1.4 Mixability and Convex Minima of Portfolios

As seen in Theorems 1.19 and 1.22, an important ingredient of the structure of worst case portfolios is played by the mixing part of this dependence structure. Next we discuss some basic results and conditions on distributions yielding n-mixing of $F_1, \dots, F_n$.

Proposition 1.24 (Mixabililty) *Let F be a distribution function on $\mathbb{R}^1$.*

a) F is 2-mixable if and only if F is symmetric, i.e., there exists $a \in \mathbb{R}$, such that $X \sim F$ implies $a - X \sim F$.
b) If F is n-mixable and T is linear, then F_T is n-mixable.
c) The binomial distribution $B(n, \frac{p}{q})$ is q-mixable.
d) The uniform distribution $U(a, b)$ on an interval (a, b) is n-mixable for any $n \geq 2$.
e) Any continuous distribution function having a symmetric unimodal density is n-mixable for any $n \geq 2$.

The statements in a), b) are obvious; for c) see Rüschendorf (1983a). The minimal variance $v_k(\vartheta)$ of $\sum_{i=1}^k X_i$, $X_i \sim \mathcal{B}(1, \vartheta)$ has been determined in Snijders (1984):

$$v_k(\vartheta) = a(k, \vartheta)(1 - a(k, \vartheta)), \quad a(k, \vartheta) = k\vartheta (\operatorname{mod} 1). \tag{1.60}$$

d) is established in Gaffke and Rüschendorf (1981), while e) is proved in Rüschendorf and Uckelmann (2002). In particular, several standard distributions, like the normal and Cauchy, are mixing. Some general mixing results are established in Wang and Wang (2011, 2016), Wang et al. (2013), and Puccetti et al. (2012). We highlight some important results from the ample theory developed in these papers.

Theorem 1.25 (Monotone densities) *Let F be a distribution function on $[a, b]$ with lower (upper) support point a (b) and mean μ.*

a) If F has an increasing density and $\mu \leq b - \frac{1}{n}(b - a)$, then F is n-mixable.
b) If F has a decreasing density and $\mu \geq a + \frac{1}{n}(b - a)$, then F is n-mixable.

The proof of Theorem 1.25 and of related mixing results is based on combinatorial approximation by discrete uniform distributions on n points, which are n-mixable. In this connection, a useful result states the convexity of the class of n-mixable distributions having the same center μ.

Proposition 1.26 (Convexity of n-mixable distributions)

a) A countable convex combination of n-mixable distribution functions with the same center μ is n-mixable.
b) A discrete distribution F is mixable with center μ if and only if it is a countable convex combination of n-discrete uniform n-mixable distributions.

This proposition leads to the following result concerning concave densities.

Theorem 1.27 (Concave densities) *Any distribution on a bounded interval (a, b) with a concave density is n-mixing for any $n \geq 3$.*

For an extension to the inhomogeneous case $F_1, \dots, F_n$, the following conditions are shown to be necessary in Wang and Wang (2016).

Proposition 1.28 (Necessary conditions for mixability) *If $F_1, \ldots, F_n$ are mixable with support (a_i, b_i) and means $\mu_1, \ldots, \mu_n$, then with the lengths $\ell_1 = b_i - a_i$ it holds that*

a) ***mean inequality:***

$$\sum_{i=1}^n a_i + \max_{1\le i\le n} \ell_i \le \sum_{i=1}^n \mu_i \le \sum_{i=1}^n b_i - \max_{1\le i\le n} \ell_i; \tag{1.61}$$

b) ***norm inequality:*** *For any p-norm, $\| \; \| = \| \; \|_p$, $p \le \infty$ and $X_i \sim F_i$ holds*

$$\sum_{i=1}^n \|X_i - \mu_i\| \ge 2 \max_{1\le i\le n} \|X_i - \mu_i\|; \tag{1.62}$$

c) ***length inequality:***

$$\sum_{i=1}^n \ell_i \ge 2 \max_{1\le i\le n} \ell_i; \tag{1.63}$$

d) ***variance inequality:***

$$\sum_{i=1}^n \sigma_i \ge 2 \max_{1\le i\le n} \sigma_i, \tag{1.64}$$

where $0 \le \sigma_i^2 \le \infty$ are the variances of F_i.

The convexity property as in Proposition 1.26 also holds in the inhomogeneous case:
If $F_1, \ldots, F_n$ and $G_1, \ldots, G_n$ are mixable, then

$$\alpha F_1 + (1-\alpha)G_1, \ldots, \alpha F_n + (1-\alpha)G_n \text{ are also mixable.} \tag{1.65}$$

These properties lead to an extension of Theorem 1.25 to the inhomogeneous case.

Theorem 1.29 (Decreasing densities) *If $F_1, \ldots, F_n$ are distributions with bounded supports and decreasing densities, satisfying the mean inequality* (1.61)*, then $F_1, \ldots, F_n$ are mixable.*

As shown in Theorem 1.15, Corollary 1.17, and Theorem 1.22, mixing allows us to determine sharp bounds for the value-at-risk and for the tail risk of the portfolio. The tail risk bounds in these results depend on the determination of the largest mixing intervals (a, b) for the tail levels allowing the construction of an admissible coupling on these intervals.

Let F be a distribution function on $[0, 1]$ with a decreasing density. For $0 \le c \le \frac{1}{n}$ we define a copula $Q_n^F(c)$ as follows:
The random variables $(U_1, \ldots, U_n)$ have distribution $Q_n^F(c)$ if

1) For each i with $U_i \in [1-c, 1]$ we have $U_j = (n-1)(1-U_i)$ for all $j \ne i$;
2) $F^{-1}(U_1) + \cdots + F^{-1}(U_n)$ is a constant when any of the $U_i \in ((n-1)c, 1-c)$,

https://doi.org/10.1017/9781009367189.004 Published online by Cambridge University Press

i.e., a large U_i is coupled with $n-1$ identical small U_j, and if one of the U_i takes values in the intermediate interval then it is mixing with the other U_j and in fact all other U_j lie in the intermediate interval. Denote

$$H(t) = (n-1)F^{-1}((n-1)t) + F^{-1}(1-t). \tag{1.66}$$

Similarly, in the case of increasing density, define $Q_n^F(c)$ by the properties:

1) For each i with $U_i \in [0, c]$ we have $U_j = 1-(n-1)U_i$, for all $j \neq i$, and
2) $G(U_1) + \cdots + G(U_n) = \text{const.}$, when any of the $U_i \in (c, 1-(n-1)c)$, $G = F^{-1}$.

In this case define

$$H(t) = G(t) + (n-1)G(1-(n-1)t). \tag{1.67}$$

Proposition 1.30

a) In the case of monotone densities and $c \in [0, \frac{1}{n}]$, there exists a copula $Q_n^F(c)$ satisfying 1), 2) if

$$\int_c^{\frac{1}{n}} H(t)\,\mathrm{d}t \leq \Big(\frac{1}{n} - c\Big)H(c). \tag{1.68}$$

b) The smallest c such that a copula $Q_n^F(c)$ with 1), 2) exists is given by

$$c_n = \min\Big\{c \in \Big[0, \frac{1}{n}\Big];\ \textit{inequality}\ (1.68)\ \textit{holds}\Big\}. \tag{1.69}$$

Proof For the proof we use that the assumed mixability for $Q_n^F(c)$ implies the moderate moment condition in Theorem 1.25 on $[c, 1-(n-1)c]$ in the increasing case resp. on $[(n-1)c, 1-c]$ in the decreasing case.

In consequence, if $Q_n^F(c)$ exists, it follows that in the increasing case the $(G(U_j))$ are mixable on $[c, 1-(n-1)c]$ when $X_i \in [c, 1-(n-1)c]$, and thus

$$\begin{aligned}\sum_{j=1}^n G(X_j) &= E\Big(\sum_{j=1}^n G(X_j) \mid c \leq U_i \leq 1-(n-1)c\Big)\\ &= \frac{n}{1-nc}\int_c^{1-(n-1)c} G(t)\,\mathrm{d}t;\end{aligned}$$

similarly in the decreasing case. By the necessary moderate moment condition in Proposition 1.28, this implies that the conditional mean is less than or equal to $\frac{G(c)}{n+n-1)}\frac{G(1-n-1)c)}{n}$, and thus

$$\int_c^{1-(n-1)c} G(t)\,\mathrm{d}t \leq \Big(\frac{1}{n} - c\Big)(G(c) + (n-1)G(1-(n-1)c).$$

This implies (1.68).

In the increasing case, $G(t)$ and thus $H(t)$ is concave on $[0, \frac{1}{n}]$. Thus the set of all c satisfying (1.68) is a closed interval $[\widehat{c}_n, \frac{1}{n}]$, and by (1.68), $\widehat{c}_n \leq c \leq \frac{1}{n}$ and thus $c_n \geq \widehat{c}_n$. Since by Proposition 1.30 a) it follows that $Q_n^F(\widehat{c}_n)$ exists, we obtain $c_n = \widehat{c}_n$. The case of decreasing densities is similar. □

https://doi.org/10.1017/9781009367189.004 Published online by Cambridge University Press

An interesting consequence of Proposition 1.30 is the following convex ordering result for the portfolios.

Theorem 1.31 (Convex minima of portfolios) *Let F have a monotone density, and let $(U_1,\dots,U_n)$ be a copula vector corresponding to $Q_n^F = Q_n^F(c_n)$; then for all $X_i \sim F$, $1 \le i \le n$, it holds that*

a)
$$\sum_{i=1}^n X_i \ge_{\mathrm{cx}} \sum_{i=1}^n G(U_i), \quad G = F^{-1}. \tag{1.70}$$

b) For any convex function f, it holds that

$$\min_{\substack{X_i\sim F_i,\\ 1\le i\le n}} Ef\Big(\sum_{i=1}^n X_i\Big) = n\int_0^{c_n} f(H(t))\,\mathrm{d}t + (1-nc_n)f(H(c_n)). \tag{1.71}$$

Proof For the proof of Theorem 1.31 a) it is established that the excess function of the right-hand side dominates that of the left-hand side (see Theorems 3.4, 3.5 in Wang and Wang, 2011), while b) is direct from construction. □

The coupling Q_n^F also solves the following minimization problem, which has an ample historical background.

Proposition 1.32 *For $(U_1,\dots,U_n) \sim Q_n^F$ it holds that*

$$\begin{aligned}\Lambda_n &= \min_{X_i\sim U(0,1)} E\prod_{i=1}^n X_i = E\prod_{i=1}^n U_i\\ &= e^{-n} + \frac{n}{2}e^{-2n} + O(n^4e^{-3n}).\end{aligned} \tag{1.72}$$

In the case of decreasing density on the support $[0,1]$ with moderate moment condition, it also implies sharpness of the upper bound for the tail risk in the conditional moment method in Theorem 1.19.

Theorem 1.33 *Let F have a decreasing density on its support $[0,1]$ with mean μ and moderate moment condition $E(X \mid X \ge t) \ge t + \frac{1-t}{n}$ for $X \sim F$ and any $t \in [0,1]$. Then for $\Psi(t) = E(X \mid X \ge G(t))$, $G = F^{-1}$, it holds that*

$$\begin{aligned}M_n^+(s) &= \sup\Big\{P\Big(\sum_{i=1}^n X_i \ge s\Big);\ \ X_i \sim F_i, 1\le i\le n\Big\}\\ &= 1 - m_n(s) = \begin{cases}1, & s \le n\mu,\\ 1-\Psi^-(\frac{s}{n}), & n\mu < s < n,\\ 0, & s \ge n.\end{cases}\end{aligned} \tag{1.73}$$

Proof By Theorem 1.19, the right-hand side of (1.73) is an upper bound for the tail risk. By Theorem 1.25, $G(V)$ is n-mixable for $V \sim U[a,1]$, $a = \Psi^-(\frac{s}{n})$. Thus there exist $V_i \sim V$ such that $\sum_{i=1}^n G(V_i) = n\Psi(a) = s$. Defining $Y_i = G(U)1_{(U\le a)} + G(V_i)1_{(U>a)}$, $U \sim U(0,1)$ independent of (V_i), we find $Y_i \sim F$ and

$$P\Big(\sum_{i=1}^n Y_i \ge s\Big) = P(U \ge a) = 1 - a = 1 - \Psi^-\Big(\frac{s}{n}\Big)$$

for $n\mu < s < n$. □

https://doi.org/10.1017/9781009367189.004 Published online by Cambridge University Press

In the non-mixable case of general support there is the following variant of Theorem 1.33.

For F with decreasing density and $a \in [0, 1]$, define modifications of H, c_n, Φ by

$$H_a(t) = (n-1)F^{-1}(a + (n-1)t) + F^{-1}(1-t), \quad t \in \Big[0, \frac{1-a}{n}\Big], \tag{1.74}$$

$$c_n(a) = \min \Big\{c \in \Big[0, \frac{1}{n}(1-a)\Big] : \tag{1.75}$$
$$\int_0^{\frac{1}{n}(1-a)} H_a(t)\,\mathrm{d}t \geq \Big(\frac{1}{n}(1-a) - c\Big)\Big) H_a(c)\Big\},$$

and $\Phi(a) = \begin{cases} H_a(c_n(a)), & \text{if } c_n(a) > 0, \\ n\Psi(a), & \text{if } c_n(a) = 0. \end{cases}$

Similarly for F with increasing density, $a \in [0, 1]$ define

$$H_a(t) = F^{-1}(a + t) + (n-1)F^{-1}(1 - (n-1)t), \tag{1.76}$$

$$c_n(a) = \min \Big\{c \in \Big[0, \frac{1}{n}(1-a)\Big] : \tag{1.77}$$
$$\int_c^{\frac{1}{n}(1-a)} H_a(t)\,\mathrm{d}t \leq \Big(\frac{1}{n}(1-a) - c\Big) H_a(c)\Big\},$$

and $\Phi(a) = \begin{cases} H_a(0), & \text{if } c_n(a) > 0 \\ n\Psi(a), & \text{if } c_n(a) = 0. \end{cases}$

With $\Phi^-(t) = 0$ if $t < \Phi(0)$ and $\Phi^-(t) = 1$ if $t > \Phi(1)$, then it holds that:

Theorem 1.34 *If F has a monotone density on its support, then the maximal tail risk $M_n^+(s)$ is given by*

$$M_n^+(s) = 1 - m_n(s) = 1 - \Phi^-(s). \tag{1.78}$$

The method of proof is similar to that of the improved extended conditional moment method in Theorem 1.22. That $1-\Phi^-(s)$ is an upper bound follows from the conditional moment method. The attainment follows by the property of an optimal coupling X that $\{S \geq s\} = \{X_i \geq F^{-1}(a)\}$, $a = m_n(s)$, implying that $m_n(s) = \Phi^-(s)$.

As a result, in the case of tail-decreasing densities, Theorem 1.34 implies a formula for the maximal tail risk for all sufficiently high risk levels s and similarly gives the worst case VaR_α-bounds for all levels $\alpha \geq \alpha_0$ that are sufficiently large.