Robustness to Dependency in Portfolio Optimization Using Overlapping Marginals

Xuan Vinh Doan *   Xiaobo Li †   Karthik Natarajan ‡

May 2013

Abstract

In this paper, we develop a distributionally robust portfolio optimization model where the robustness is to different dependency structures among the random losses. For a Fréchet class of distributions with overlapping marginals, we show that the distributionally robust portfolio optimization problem is efficiently solvable with linear programming. To guarantee the existence of a joint multivariate distribution consistent with the overlapping marginal information, we make use of the graph theoretic - running intersection property. We use this property to develop a tight linear programming formulation. Lastly, we use a data-driven approach using real financial data to identify the Fréchet class of distributions with overlapping marginals and then optimize the portfolio over this class of distributions. Our results show that the optimization models proposed in this paper improves on the sample based approach.

1 Introduction

Optimization under uncertainty is an active research area with many interesting applications in the area of risk management. An example of a risk management problem is to choose a portfolio

*DIMAP and ORMS Group, Warwick Business School, University of Warwick, Coventry, CV4 7AL, United Kingdom, xuan.doan@wbs.ac.uk.

†Engineering Systems and Design, Singapore University of Technology and Design, 20 Dover Drive, Singapore 138682, li_xiaobo@sutd.edu.sg

‡Engineering Systems and Design, Singapore University of Technology and Design, 20 Dover Drive, Singapore 138682, natarajan_karthik@sutd.edu.sg

of assets such that the joint portfolio risk is minimized while a certain level of the expected return is guaranteed. Markowitz [25] was the first to model this problem using variance as the risk measure. Several alternative risk measures have been proposed since for this problem. Value-at-risk (VaR) and conditional value-at-risk (CVaR) are two such popular risk measures (see, for example, Jorion [16] and Rockafellar and Uryasev [33]). However even assuming that the joint distribution of the random losses is known, the calculation of VaR and CVaR for a given portfolio involves multidimensional integrals, which can be computationally challenging and one often resorts to simulation. For discrete distributions where the losses of the different assets are independent, the computation of VaR and CVaR requires the consideration of all the support points, which can be exponentially large compared to the marginal support points. Furthermore, if the assumed joint distribution does not match the actual distribution, the optimal solution obtained by this model might perform poorly in the out of sample data.

One popular approach to address this issue is that instead of assuming a complete joint distribution for the random losses of the risky assets, one only assumes partial distributional information. Given the partial distributional information, it is natural to compute the worst case bounds for the VaR and CVaR measures. Several alternative models have been used to capture the ambiguity in distributions. This includes the simple Fréchet class of distributions in which only the univariate marginal distributions are specified (see Meilijson and Nadas [26]) and the class of distributions where multivariate marginals of non-overlapping subsets of assets are specified (see Doan and Natarajan [4], Garlappi, Uppal and Wang [21]). In this paper, we adopt a more general representation of distributional uncertainty where the multivariate marginals possibly overlap with each other.

1.1 Fréchet Class of Distributions

Let $\tilde{\boldsymbol{c}}$ be a N-dimensional random vector and θ denote the joint distribution of $\tilde{\boldsymbol{c}}$. Consider the evaluation of the expected value of a convex piecewise linear function constructed by taking the maximum of M affine functions:

$$\mathbb{E}_\theta\left[\max_{j=1,..,M}\ \left(\boldsymbol{a_j}^T\tilde{\boldsymbol{c}}+b_j\right)\right]. \tag{1}$$

Let $\mathcal{M} = \{1, \ldots, M\}$ index the affine functions in $\varphi_N(\boldsymbol{c}) \triangleq \max_{j \in \mathcal{M}} \boldsymbol{a_j}^T \boldsymbol{c} + b_j$. Associated to this is a stochastic optimization problem of the form:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}_\theta \left[\max_{j \in \mathcal{M}} \left(\boldsymbol{a_j}(\boldsymbol{x})^T \tilde{\boldsymbol{c}} + b_j(\boldsymbol{x}) \right) \right], \tag{2}$$

where $\boldsymbol{a_j}$ and b_j are affine functions of a decision vector $\boldsymbol{x}$ that is assumed to lie in a feasible region $\mathcal{X}$. Portfolio optimization with the CVaR measure lies within the scope of problem (2). Let $\tilde{c}_i$ represent the random loss of the ith asset in a financial portfolio. The total portfolio loss is then $\sum_{i=1}^{N} \tilde{c}_i x_i$, where x_i is the allocation in the ith asset. The CVaR of the portfolio loss for a given risk level $\alpha \in (0, 1)$ is defined as (see Rockafellar and Uryasev [33, 34]):

$$\text{CVaR}_\alpha^\theta(\boldsymbol{x}) \triangleq \min_{\beta \in \Re} \left\{ \beta + \frac{1}{1-\alpha} \mathbb{E}_\theta \left[\left(\sum_{i \in \mathcal{N}} \tilde{c}_i x_i - \beta \right)^+ \right] \right\}, \quad \alpha \in (0, 1),$$

where $x^+ = \max\{0, x\}$. The portfolio optimization problem is then formulated as:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \text{CVaR}_\alpha^\theta(\boldsymbol{x}) = \min_{\beta \in \Re, \boldsymbol{x} \in \mathcal{X}} \left\{ \beta + \frac{1}{1-\alpha} \mathbb{E}_\theta \left[\left(\sum_{i \in \mathcal{N}} \tilde{c}_i x_i - \beta \right)^+ \right] \right\},$$

which is an instance of problem (2). In the distributionally robust optimization setting, the joint distribution θ is not known exactly and assumed to belong to a class Θ of distributions. Then, it is possible to calculate tight bounds of the form:

$$M(\varphi_N) = \sup_{\theta \in \Theta} \mathbb{E}_\theta \left[\max_{j \in \mathcal{M}} \left(\boldsymbol{a_j}^T \tilde{\boldsymbol{c}} + b_j \right) \right], \quad m(\varphi_N) = \inf_{\theta \in \Theta} \mathbb{E}_\theta \left[\max_{j \in \mathcal{M}} \left(\boldsymbol{a_j}^T \tilde{\boldsymbol{c}} + b_j \right) \right]. \tag{3}$$

With reference to the stochastic optimization problem, we focus only on the upper bound of a convex function in this paper since this is used to formulate the distributionally robust counterpart of (2):

$$\min_{\boldsymbol{x} \in \mathcal{X}} \sup_{\theta \in \Theta} \mathbb{E}_\theta \left[\max_{j \in \mathcal{M}} \left(\boldsymbol{a_j}(\boldsymbol{x})^T \tilde{\boldsymbol{c}} + b_j(\boldsymbol{x}) \right) \right]. \tag{4}$$

Fréchet studied bounds on the cumulative distribution function of a random vector given univariate marginal distributions (Fréchet [9], [10]). This class of distributions is referred to as the Fréchet class of distributions. Let $\mathcal{N} = \{1, \ldots, N\}$ denote the set of indices of the random vector and let $\mathcal{E} = \{J_1, ..., J_R\} \subseteq 2^{\mathcal{N}}$ be a *cover* of $\mathcal{N}$, i.e., $\cup_{r \in \mathcal{R}} J_r = \mathcal{N}$, where $\mathcal{R} = \{1, \ldots, R\}$. Assume that there is no inclusion among subsets, i.e., $J_r \nsubseteq J_s$ for any $r \neq s$. Typical covers include partitions, star covers, and series covers:

- **Partition:** $\mathcal{E}$ is a partition (non-overlapping cover) if for all $r \neq s$, $J_r \cap J_s = \emptyset$. The partition:

  $$\mathcal{E} = \{\{1\}, \{2\}, \ldots, \{N\}\},$$

  is called the *simple partition* and is the basic Fréchet class of distributions. When the subsets consist of more than one random variable, the partition is a *non-overlapping multivariate marginal cover.*

- **Star cover:** Let $\mathcal{E}_0 = \{I_0, I_1, ..., I_R\}$ be a partition of $\mathcal{N}$. Then $\mathcal{E}$ is a star cover if $J_r = I_r \cup I_0$ for all $r \in \mathcal{R}$. The star cover:

  $$\mathcal{E} = \{\{1,2\}, \{1,3\}, \ldots, \{1,N\}\},$$

  is called the *simple star cover.*

- **Series cover:** Let $\mathcal{E}_0 = \{I_0, I_1, ..., I_R\}$ be a partition of $\mathcal{N}$. Then $\mathcal{E}$ is a series cover if $J_r = I_{r-1} \cup I_r$ for all $r \in \mathcal{R}$. The series cover:

  $$\mathcal{E} = \{\{1,2\}, \{2,3\}, \ldots, \{N-1,N\}\},$$

  is called the *simple series cover.*

Given a joint distribution θ of the random vector $\tilde{\boldsymbol{c}}$, let $\mathrm{proj}_{J_r}(\theta)$ denote the multivariate marginal distribution of the random sub-vector $\tilde{\boldsymbol{c}}_r \triangleq \{\tilde{c}_i\}_{i \in J_r}$. The Fréchet class of distributions $\Theta_{\mathcal{E}}$ is defined by a set of fixed multivariate marginal distributions $\{\theta_r\}_{r \in \mathcal{R}}$ of the random sub-vectors $\tilde{\boldsymbol{c}}_r$ as follows:

$$\Theta_{\mathcal{E}} \triangleq \{\theta \mid \mathrm{proj}_{J_r}(\theta) = \theta_r, \ \forall r \in \mathcal{R}\}.$$

Given a real-valued function $\varphi(\cdot)$, the Fréchet bounds are defined as

$$M_{\mathcal{E}}(\varphi) = \sup_{\theta \in \Theta_{\mathcal{E}}} \mathbb{E}_\theta \left[\varphi(\tilde{\boldsymbol{c}})\right], \quad m_{\mathcal{E}}(\varphi) = \inf_{\theta \in \Theta_{\mathcal{E}}} \mathbb{E}_\theta \left[\varphi(\tilde{\boldsymbol{c}})\right]. \tag{5}$$

Several previous studies have focused on Fréchet bounds for the cumulative distribution function of the sum of dependent risks, i.e., $\varphi(\tilde{\boldsymbol{c}}) = \varphi_d^-(\tilde{\boldsymbol{c}}; x) = \mathbb{I}\left\{\sum_{i \in \mathcal{N}} \tilde{c}_i < x\right\}$, where $\mathbb{I}$ is an indicator function. Fréchet bounds for φ_d^- were first developed by Makarov [24] and Rüschendorf [36] for $N = 2$ given a simple partition. For $N \geq 3$, Kreinovich and Ferson [20] showed that computing

$m_{\mathcal{E}}(\varphi_d^-)$, where $\mathcal{E}$ is the simple partition, is already NP-hard. Several lower bounds of $m_{\mathcal{E}}(\varphi_d^-)$ have been proposed, among which is the *standard bound* of Embrechts, Höing and Juri [6], Rüschendorf [39], and Embrechts and Puccetti [7]:

$$\inf_{\theta \in \Theta_{\{\{1\},\ldots,\{N\}\}}} \mathbb{P}_\theta \left(\sum_{i \in \mathcal{N}} \tilde{c}_i < x \right) \geq \max \left\{ \sup_{\boldsymbol{d}: \sum_{i \in \mathcal{N}} d_i = x} \left[F_1^-(d_1) + \sum_{i=2}^{N} F_i(d_i) \right] - (N-1), 0 \right\}, \tag{6}$$

where $F_i(d_i) = \mathbb{P}(\tilde{c}_i \leq d_i)$ and $F_i^-(d_i) = \mathbb{P}(\tilde{c}_i < d_i)$ for all $i \in \mathcal{N}$. The standard bound is tight when $N = 2$. This result can be directly translated to an upper bound on the tail probability $\mathbb{P}_\theta\left(\sum_{i \in \mathcal{N}} \tilde{c}_i \geq x\right)$ given univariate marginals. Wang and Wang [42] and Pucetti and Rüschendorf [32] developed tight bounds for homogeneous univariate marginals with monotone densities and concave densities respectively. For general non-overlapping covers, Puccetti and Rüschendorf [31] showed that Fréchet bounds for φ_d^- can be reduced to that of a simple partition. For the simple star cover, Rüschendorf [37] introduced a conditioning method to derive Fréchet bounds using the results for the simple partition. For the simple series cover, Embrechts and Puccetti [8] proposed a variable splitting method to estimate Fréchet bounds from those with simple partition. Puccetti and Rüschendorf [31] generalized this method for overlapping covers. In general, given the hardness of computing the tight bounds, these methods typically generate lower bounds for $m_{\mathcal{E}}(\varphi_d^-)$ which are not tight.

In this paper, we compute the Fréchet upper bound for a convex piecewise linear function and solve the associated distributionally robust optimization problem. In Section 2, we provide conditions under which this bound is efficiently computable and use it to estimate Fréchet bounds on the cumulative distribution function of the sum of dependent risks in Section 3. In Section 4, we propose a data-driven approach to estimate the cover structure of the Fréchet class of distributions and then optimize the portfolio returns using real financial data before concluding in Section 5.

2 Fréchet Bounds with Regular Covers

For a partition (non-overlapping cover), the Fréchet class of distributions $\Theta_{\mathcal{E}}$ is guaranteed to be non-empty given any feasible set of multivariate marginal distributions $\{\theta_r\}_{r \in \mathcal{R}}$. Feasibility is ensured using a product measure on the multivariate marginal distributions. However for arbitrary covers with overlaps, the feasibility problem is itself non-trivial. For an overlapping cover, $\Theta_{\mathcal{E}} \neq$

$\emptyset$ implies the *consistency* of the multivariate marginals, namely, $\text{proj}_{J_r \cap J_s}(\theta_r) \equiv \text{proj}_{J_r \cap J_s}(\theta_s)$ for all $r \neq s$. The reverse need not be true however, that is, consistency does not necessarily imply that $\Theta_{\mathcal{E}} \neq \emptyset$. Vorob'ev [40] provided a simple counterexample with three variables $\mathcal{E} = \{\{1,2\},\{2,3\},\{1,3\}\}$ where $\mathbb{P}(\tilde{c}_1 = 1, \tilde{c}_2 = 1) = \mathbb{P}(\tilde{c}_1 = 0, \tilde{c}_2 = 0) = 1/2$, $\mathbb{P}(\tilde{c}_1 = 1, \tilde{c}_3 = 1) = \mathbb{P}(\tilde{c}_1 = 0, \tilde{c}_3 = 0) = 1/2$ and $\mathbb{P}(\tilde{c}_2 = 1, \tilde{c}_3 = 0) = \mathbb{P}(\tilde{c}_2 = 0, \tilde{c}_3 = 1) = 1/2$. In this case, $\Theta_{\mathcal{E}} = \emptyset$ even though the given marginal distributions are consistent. A cover for which consistency is sufficient to ensure $\Theta_{\mathcal{E}}$ is non-empty is termed a *regular cover* (see Vorob'ev [40]). If the cover is *irregular*, there is no simple sufficient condition to check if $\Theta_{\mathcal{E}}$ is empty or not (see Kellerer [19]).

2.1 Regular Covers

In this section, we study the computation of the Fréchet upper bound $M_{\mathcal{E}}(\varphi_N)$ for regular covers. Regularity is verifiable by an equivalent graph theoretic property of the cover known as the *running intersection property* (RIP) (see Lauritzen et al. [23]). A cover $\mathcal{E}$ satisfies the RIP, if the elements of $\mathcal{E}$ can be ordered in such a way that:

$$\forall r \in \mathcal{R} \setminus \{1\},\ \exists \sigma_r \in \mathcal{R} \text{ s.t. } 1 \leq \sigma_r < r \text{ and } J_r \cap \left(\bigcup_{t=1}^{r-1} J_t \right) \subseteq J_{\sigma_r}. \tag{7}$$

Then, $\mathcal{E}$ is regular. The star cover is a regular cover with $\sigma_r = 1$. Similarly, the series cover is a regular cover with $\sigma_r = r-1$. From this point onwards, we assume that the elements of the cover are ordered as per (7). The sparse structure of the cover implied by the running intersection property has been used in developing tractable semidefinite relaxations for sparse polynomial optimization problems by Lasserre [22] and in developing efficient inference algorithms for probabilistic graphical models (see Wainwright [41]). Define the parameters:

$$\begin{aligned}
\sigma_r &\triangleq \min \left\{ i \in \mathcal{R} \;\middle|\; J_r \cap \left(\bigcup_{t=1}^{r-1} J_t \right) \subseteq J_i \right\}, && \forall r \in \mathcal{R} \setminus \{1\}, \\
K_r &\triangleq J_r \cap \left(\bigcup_{t=1}^{r-1} J_t \right), && \forall r \in \mathcal{R} \setminus \{1\}.
\end{aligned}$$

Given consistent marginals and a regular cover, a feasible joint distribution is constructed using conditional independence as follows (see Kellerer [17] and Jiroušek [15]):

$$\theta(\boldsymbol{c}) = \theta_1(\boldsymbol{c}_1) \times \frac{\theta_2(\boldsymbol{c}_2)}{\text{proj}_{K_2}(\theta_2(\boldsymbol{c}_2))} \times \cdots \times \frac{\theta_R(\boldsymbol{c}_R)}{\text{proj}_{K_R}(\theta_R(\boldsymbol{c}_R))}, \quad \forall \boldsymbol{c}. \tag{8}$$

In the simple star case, this reduces to:

$$\theta(\boldsymbol{c}) = \theta_1(c_1) \times \theta_2(c_2|c_1) \times \theta_3(c_3|c_1) \cdots \times \theta_R(c_R|c_1), \quad \forall \boldsymbol{c}, \tag{9}$$

where $\theta_r(c_r|c_1)$ is the conditional distribution of c_r given c_1 while in the simple series case, this reduces to:

$$\theta(\boldsymbol{c}) = \theta_1(c_1) \times \theta_2(c_2|c_1) \times \theta_3(c_3|c_2) \times \cdots \times \theta_R(c_R|c_{R-1}), \quad \forall \boldsymbol{c}. \tag{10}$$

Define the vector $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)$ with $\eta_i = \left(\sum_{r \in \mathcal{R}} \mathbb{I}\{i \in J_r\}\right)^{-1}$ for all $i \in \mathcal{N}$. The split vector $\boldsymbol{\eta}$ allows us to reexpress the linear function in $\tilde{\boldsymbol{c}}$ as separable functions with respect to the cover $\mathcal{E}$. In particular, $\tilde{\boldsymbol{c}}^T \boldsymbol{a} = \sum_{r \in \mathcal{R}} \tilde{\boldsymbol{c}}_r^T (\boldsymbol{\eta}_r \circ \boldsymbol{a}_r)$ where $\circ$ is the Hadamard (entry-wise) product operator. Finally, let B_r denote the support of $\tilde{\boldsymbol{c}}_r$ for all $r \in \mathcal{R}$. Our first theorem provides an infinite-dimensional linear optimization formulation to compute the Fréchet upper bound.

Theorem 1 *Given a regular cover $\mathcal{E}$ and a consistent set $\{\theta_r\}_{r \in \mathcal{R}}$ of marginal distributions with finite second moments, define $M_{\mathcal{E}}^{\mathrm{P}}(\varphi_N)$ as the optimal value to the problem:*

$$\begin{aligned}
M_{\mathcal{E}}^{\mathrm{P}}(\varphi_N) = \sup_{\vartheta_{j,r}(\boldsymbol{c}_r), \lambda_j} \quad & \sum_{j \in \mathcal{M}} \sum_{r \in \mathcal{R}} \int_{B_r} \boldsymbol{c}_r^T (\boldsymbol{\eta}_r \circ \boldsymbol{a_j}_r) \cdot \vartheta_{j,r}(\boldsymbol{c}_r) \ \mathrm{d}\boldsymbol{c}_r + \sum_{j \in \mathcal{M}} b_j \lambda_j \\
\text{s.t.} \quad & \vartheta_{j,r}(\boldsymbol{c}_r) \geq 0, && \forall r \in \mathcal{R}, \forall \boldsymbol{c}_r \in B_r, \forall j \in \mathcal{M}, \\
& \sum_{j \in \mathcal{M}} \vartheta_{j,r}(\boldsymbol{c}_r) = \theta_r(\boldsymbol{c}_r), && \forall r \in \mathcal{R}, \forall \boldsymbol{c}_r \in B_r, \\
& \int_{B_r} \vartheta_{j,r}(\boldsymbol{c}_r) \ \mathrm{d}\boldsymbol{c}_r = \lambda_j, && \forall r \in \mathcal{R}, \forall j \in \mathcal{M}, \\
& proj_{K_r}(\vartheta_{j,r}(\boldsymbol{c}_{K_r})) = proj_{K_r}(\vartheta_{j,\sigma_r}(\boldsymbol{c}_{K_r})), && \forall r \in \mathcal{R} \setminus \{1\} : K_r \neq \emptyset, \\
& && \forall \boldsymbol{c}_{K_r} \in B_{K_r}, \forall j \in \mathcal{M}.
\end{aligned} \tag{11}$$

Then the Fréchet bound $M_{\mathcal{E}}(\varphi_N) = \sup_{\theta \in \Theta_{\mathcal{E}}} \mathbb{E}_\theta \left[\max_{j \in \mathcal{M}} \left(\boldsymbol{a_j}^T \tilde{\boldsymbol{c}} + b_j\right)\right]$ is equal to $M_{\mathcal{E}}^{\mathrm{P}}(\varphi_N)$.

To prove this theorem, we need the following lemma regarding the consistency of marginal distributions.

Lemma 1 *Given a regular cover $\mathcal{E}$, the following condition is necessary and sufficient to ensure consistency among the marginal distributions:*

$$proj_{K_r}(\theta_r) = proj_{K_r}(P_{\sigma_r}), \ \forall r \in \mathcal{R} \setminus \{1\} : K_r \neq \emptyset$$

Proof. Using the RIP condition in (7), we have for all $r \in \mathcal{R} \setminus \{1\}$,

$$
\begin{aligned}
J_r \cap \left(\overset{r-1}{\underset{t=1}{\cup}} J_t\right) \subseteq J_{\sigma_r} \quad &\Longleftrightarrow \quad J_r \cap \left(\overset{r-1}{\underset{t=1}{\cup}} J_t\right) \subseteq J_{\sigma_r} \cap J_r, \\
&\Longleftrightarrow \quad J_r \cap J_t \subseteq J_{\sigma_r} \cap J_r, \qquad \forall t = 1, \ldots, r-1.
\end{aligned}
$$

This indicates that all the pairwise intersections are included in a set of $N-1$ intersections. Thus verifying consistency can be restricted to these pairs. Since, $J_{\sigma_r} \cap J_r = \left(\overset{r-1}{\underset{t=1}{\cup}} J_t\right) \cap J_r = K_r$, this implies the result. □

We now turn to the proof of Theorem 1.

Proof. We first show that $M_{\mathcal{E}}^{\mathrm{P}}(\varphi_N)$ is a valid upper bound of $M_{\mathcal{E}}(\varphi_N)$. For any joint distribution $\theta \in \Theta_{\mathcal{E}}$, the expected value in (1) can be expressed as follows:

$$
\begin{aligned}
\mathbb{E}_\theta\left[\varphi_N(\tilde{\boldsymbol{c}})\right] &= \mathbb{E}_\theta\left[\max_{j \in \mathcal{M}}\left(\sum_{r \in \mathcal{R}} \tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r \circ \boldsymbol{a}_{\boldsymbol{j}_r}) + b_j\right)\right] \\
&= \sum_{j \in \mathcal{M}} \mathbb{E}_\theta\left[\sum_{r \in \mathcal{R}} \tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r \circ \boldsymbol{a}_{\boldsymbol{j}_r}) + b_j \;\middle|\; \text{the } j\text{th term is max}\right] \mathbb{P}_\theta\left(\text{the } j\text{th term is max}\right) \\
&= \sum_{j \in \mathcal{M}}\left(\sum_{r \in \mathcal{R}} \mathbb{E}_\theta\left[\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r \circ \boldsymbol{a}_{\boldsymbol{j}_r}),\ \text{the } j\text{th term is max}\right] + b_j \lambda_j\right) \\
&= \sum_{j \in \mathcal{M}} \sum_{r \in \mathcal{R}} \int_{B_r} \boldsymbol{c}_r^T(\boldsymbol{\eta}_r \circ \boldsymbol{a}_{\boldsymbol{j}_r}) \cdot \mathbb{P}_\theta\left(\tilde{\boldsymbol{c}}_r = \boldsymbol{c}_r,\ \text{the } j\text{th term is max}\right)\, \mathrm{d}\boldsymbol{c}_r \\
&\quad + \sum_{j \in \mathcal{M}} b_j \mathbb{P}_\theta\left(\text{the } j\text{th term is max}\right) \\
&= \sum_{j \in \mathcal{M}} \sum_{r \in \mathcal{R}} \int_{B_r} \boldsymbol{c}_r^T(\boldsymbol{\eta}_r \circ \boldsymbol{a}_{\boldsymbol{j}_r}) \cdot \vartheta_{j,r}(\boldsymbol{c}_r)\, \mathrm{d}\boldsymbol{c}_r + \sum_{j \in \mathcal{M}} b_j \lambda_j.
\end{aligned}
$$

The decision variables in this formulation are the measures $\vartheta_{j,r}(\boldsymbol{c}_r)$ for $j \in \mathcal{M}$ and $r \in \mathcal{R}$ and the probability λ_j for $j \in \mathcal{M}$:

$$
\begin{aligned}
\vartheta_{j,r}(\boldsymbol{c}_r) &\triangleq \mathbb{P}_\theta\left(\tilde{\boldsymbol{c}}_r = \boldsymbol{c}_r,\ \text{the } j\text{th term is max}\right), \text{ and} \\
\lambda_j &\triangleq \mathbb{P}_\theta\left(\text{the } j\text{th term is max}\right).
\end{aligned}
$$

Thus $\vartheta_{j,r}(\boldsymbol{c}_r) \geq 0$ for all $r \in \mathcal{R}$, $\boldsymbol{c}_r \in B_r$, and $j \in \mathcal{M}$. Note that if the function value $\varphi_N(\boldsymbol{c})$ has multiple terms attaining the maximum for some value of $\boldsymbol{c}$, one can arbitrarily choose any one of them without changing the expected value, for example, the term with the minimum index. Hence, for all $r \in \mathcal{R}$ and $\boldsymbol{c}_r \in B_r$,

$$
\begin{aligned}
\sum_{j \in \mathcal{M}} \vartheta_{j,r}(\boldsymbol{c}_r) &= \mathbb{P}_\theta\left(\tilde{\boldsymbol{c}}_r = \boldsymbol{c}_r\right) \\
&= \theta_r(\boldsymbol{c}_r).
\end{aligned}
$$

In addition, for all $r \in \mathcal{R}$ and $j \in \mathcal{M}$,

$$\begin{aligned}\int_{B_r} \vartheta_{j,r}(\boldsymbol{c}_r)\ \mathrm{d}\boldsymbol{c}_r &= \mathbb{P}_\theta\left(\text{the } j\text{th term is max}\right) \\ &= \lambda_j.\end{aligned}$$

Applying Lemma 1, the last set of constraints are necessary constraints for consistency. Thus, for any distribution $\theta \in \Theta_{\mathcal{E}}$, all the constraints in (11) are satisfied, which implies $M_{\mathcal{E}}^{\mathrm{P}}(\varphi_N)$ is an upper bound of $\mathbb{E}_\theta\left[\varphi_N(\tilde{\boldsymbol{c}})\right]$. We then have

$$M_{\mathcal{E}}^{\mathrm{P}}(\varphi_N) \geq M_{\mathcal{E}}(\varphi_N) = \sup_{\theta\in\Theta_{\mathcal{E}}} \mathbb{E}_\theta\left[\varphi_N(\tilde{\boldsymbol{c}})\right].$$

We next prove that the bound is tight. Since all marginal distributions θ_r, $r \in \mathcal{R}$, have finite second moments, problem (11) is bounded. Consider an optimal solution of problem (11) denoted by $\vartheta_{j,r}^*(\boldsymbol{c}_r)$ and λ_j^*, $j \in \mathcal{M}$, $r \in \mathcal{R}$, $\boldsymbol{c}_r \in B_r$. We have:

$$\begin{aligned}\sum_{j\in\mathcal{M}} \lambda_j^* &= \sum_{j\in\mathcal{M}} \int_{B_r} \vartheta_{j,r}^*(\boldsymbol{c}_r)\ \mathrm{d}\boldsymbol{c}_r \\ &= \int_{B_r} \theta_r(\boldsymbol{c}_r)\ \mathrm{d}\boldsymbol{c}_r \\ &= 1.\end{aligned}$$

In addition, $\lambda_j^* \geq 0$ for all $j \in \mathcal{M}$. Thus $\boldsymbol{\lambda}^*$ is a probability vector. We now construct a distribution θ^* for $\tilde{\boldsymbol{c}}$ based on $\vartheta_{j,r}^*(\boldsymbol{c}_r)$ and λ_j^* as follows:

(a) Choose term j with probability λ_j^* for $j \in \mathcal{M}$.

(b) For each $r \in \mathcal{R}$, assign a measure $\theta_{j,r}^*(\boldsymbol{c}_r)$ for $\boldsymbol{c}_r \in B_r$ where $\theta_{j,r}^*(\boldsymbol{c}_r) = \vartheta_{j,r}^*(\boldsymbol{c}_r)/\lambda_j^*$.

(c) Choose a feasible joint distribution in the Fréchet class of distributions $\theta_j^* \in \Theta_{\mathcal{E}}(\theta_{j,r}^*, r \in \mathcal{R})$ and generate $\tilde{\boldsymbol{c}}$ with distribution θ_j^*.

We assume that $\lambda_j^* > 0$. If there exists an index with probability 0, we drop that index. It is clear that $\theta_{j,r}^*$ is a valid and consistent probability measure for $\tilde{\boldsymbol{c}}_r$, $r \in \mathcal{R}$, since $(\vartheta_{j,r}^*(\boldsymbol{c}_r), \lambda_j^*)$ is a feasible solution to problem (11). Thus, $\Theta_{\mathcal{E}}(\theta_{j,r}^*, r \in \mathcal{R}) \neq \emptyset$, which implies the existence of a θ_j^* for all $j \in \mathcal{M}$. For all $r \in \mathcal{R}$, the joint probability of $\tilde{\boldsymbol{c}}_r$ taking the $\boldsymbol{c}_r$ value is:

$$\begin{aligned}\sum_{j\in\mathcal{M}} \lambda_j^* \cdot \frac{\vartheta_{j,r}^*(\boldsymbol{c}_r)}{\lambda_j^*} &= \sum_{j\in\mathcal{M}} \vartheta_{j,r}^*(\boldsymbol{c}_r) \\ &= \theta_r(\boldsymbol{c}_r).\end{aligned}$$

Thus, we have $\theta^* \in \Theta_{\mathcal{E}}$. Then the following inequality holds:

$$\begin{aligned}
\mathbb{E}_{\theta_j^*}\left[\max_{k\in\mathcal{M}}\left(\sum_{r\in\mathcal{R}}\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{\boldsymbol{k}r})+b_k\right)\right] &\geq \mathbb{E}_{\theta_j^*}\left[\sum_{r\in\mathcal{R}}\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{\boldsymbol{j}r})+b_j\right] \\
&= \sum_{r\in\mathcal{R}}\mathbb{E}_{\theta_j^*}\left[\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{\boldsymbol{j}r})+b_j\right] \\
&= \sum_{r\in\mathcal{R}}\frac{1}{\lambda_j^*}\int_{B_r}\boldsymbol{c}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{\boldsymbol{j}r})\cdot\vartheta_{j,r}^*(\boldsymbol{c}_r)\ \mathrm{d}\boldsymbol{c}_r+b_j,
\end{aligned}$$

where the first inequality is obtained by simple choosing the jth term in the function for θ_j^*. Then we have a lower bound since:

$$\begin{aligned}
\mathbb{E}_{\theta^*}\left[\max_{k\in\mathcal{M}}\left(\sum_{r\in\mathcal{R}}\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{\boldsymbol{k}r})+b_k\right)\right] &= \sum_{j\in\mathcal{M}}\lambda_j^*\cdot\mathbb{E}_{\theta_j^*}\left[\max_{k\in\mathcal{M}}\left(\sum_{r\in\mathcal{R}}\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{\boldsymbol{k}r})+b_k\right)\right] \\
&\geq \sum_{j\in\mathcal{M}}\lambda_j^*\left[\sum_{r\in\mathcal{R}}\frac{1}{\lambda_j^*}\int_{B_r}\boldsymbol{c}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{\boldsymbol{j}r})\cdot\vartheta_{j,r}^*(\boldsymbol{c}_r)\ \mathrm{d}\boldsymbol{c}_r+b_j\right] \\
&= \sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{M}}\int_{B_r}\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{\boldsymbol{j}r})\cdot\vartheta_{j,r}^*(\boldsymbol{c}_r)\ \mathrm{d}\boldsymbol{c}_r+b_j\lambda_j^* \\
&= M_{\mathcal{E}}^P(\varphi_N).
\end{aligned}$$

Hence

$$M_{\mathcal{E}}(\varphi_N)=\sup_{\theta\in\Theta_{\mathcal{E}}}\ \mathbb{E}_\theta\left[\max_{j\in\mathcal{M}}\ \left(\boldsymbol{a_j}^T\tilde{\boldsymbol{c}}+b_j\right)\right]\geq\mathbb{E}_{\theta^*}\left[\max_{j\in\mathcal{M}}\ \left(\boldsymbol{a_j}^T\tilde{\boldsymbol{c}}+b_j\right)\right]\geq M_{\mathcal{E}}^P(\varphi_N).$$

Together, we have $M_{\mathcal{E}}(\varphi_N)=M_{\mathcal{E}}^{\mathrm{P}}(\varphi_N)$. □

Several remarks about this theorem are in order:

(a) The proof of Theorem 1 is inspired from the proofs in Bertsimas, Natarajan and Teo [3] and Natarajan, Song and Teo [30] for univariate marginals and Doan and Natarajan [4] for non-overlapping multivariate marginals. We extend these results to overlapping multivariate marginals using the running intersection property.

(b) The last set of constraints in (11) ensures the consistency and hence the feasibility of a joint distribution given an overlapping regular cover. This worst-case distribution is different from the joint distribution constructed in (8). Particularly, the conditionally independent distribution in (8) is in fact the distribution that maximizes the Shannon entropy among all the measures $\theta\in\Theta_{\mathcal{E}}$ (Jiroušek [15]). Theorem 1 provides an alternate distribution in the set $\Theta_{\mathcal{E}}$ that maximizes the expected value of a piecewise linear convex function.

(c) The representation of the split vector $\boldsymbol{\eta}$ is not unique. In particular, instead of single values η_i, $i \in \mathcal{N}$, for all $r \in \mathcal{R}$, we can define $\eta_i^r \geq 0$, $i \in \mathcal{N}$ and $r \in \mathcal{R}$, such that $\sum_{r \in \mathcal{R}} \eta_i^r = 1$ for all $i \in \mathcal{N}$, and $\eta_i^r = 0$ if $i \notin J_r$ for all $r \in \mathcal{R}$ and $i \in \mathcal{N}$. For example, instead of splitting variables equally among all relevant subsets as in the current setting, we can set $\eta_i^{r(i)} = 1$ for all $i \in \mathcal{N}$, where $r(i) = \min\{r : i \in J_r\}$. This does not affect the result of Theorem 1.

(d) Suppose θ_r, $r \in \mathcal{R}$, are discrete distributions. Then the problem (11) becomes a standard finite-dimensional, linear program.

$$
\begin{aligned}
M_{\mathcal{E}}^{\mathrm{P}}(\varphi_N) = \sup_{\vartheta_{j,r}(\boldsymbol{c}_r),\lambda_j} \quad & \sum_{j\in\mathcal{M}} \sum_{r\in\mathcal{R}} \sum_{\boldsymbol{c}_r \in B_r} \boldsymbol{c}_r^T(\boldsymbol{\eta}_r \circ \boldsymbol{a}_{\boldsymbol{j}_r}) \cdot \vartheta_{j,r}(\boldsymbol{c}_r) + \sum_{j\in\mathcal{M}} b_j \lambda_j \\
\text{s.t.} \quad & \vartheta_{j,r}(\boldsymbol{c}_r) \geq 0, && \forall r \in \mathcal{R}, \forall \boldsymbol{c}_r \in B_r, \forall j \in \mathcal{M}, \\
& \sum_{j\in\mathcal{M}} \vartheta_{j,r}(\boldsymbol{c}_r) = \theta_r(\boldsymbol{c}_r), && \forall r \in \mathcal{R}, \forall \boldsymbol{c}_r \in B_r, \\
& \sum_{\boldsymbol{c}_r \in B_r} \vartheta_{j,r}(\boldsymbol{c}_r) = \lambda_j, && \forall r \in \mathcal{R}, \forall j \in \mathcal{M}, \\
& \mathrm{proj}_{K_r}(\vartheta_{j,r}(\boldsymbol{c}_{K_r})) = \mathrm{proj}_{K_r}(\vartheta_{j,\sigma_r}(\boldsymbol{c}_{K_r})), && \forall r \in \mathcal{R} \setminus \{1\} : K_r \neq \emptyset, \\
& && \forall \boldsymbol{c}_{K_r} \in B_{K_r}, \forall j \in \mathcal{M}.
\end{aligned} \tag{12}
$$

The total number of decision variables in this formulation is $M \sum_{r\in\mathcal{R}} |B_r| + M$, while the total number of constraints is $M \sum_{r\in\mathcal{R}} |B_r| + RM + \sum_{r\in\mathcal{R}} |B_r| + M \sum_{r\in\mathcal{R}\setminus\{1\}} |B_{K_r}|$. When the number of marginal supports points of each variable $\tilde{c}_i$ is bounded from above by a constant say K, and the size of each subset J_r is $O(\log(N))$, then the size of the set $|B_r| \leq K^{O(\log(N))} = poly(N)$. Problem (11) can then be solved in polynomial time in N, M, R and K.

We conclude this section by showing that the result in Doan and Natarajan [4] for general partitions can be derived from the result of Theorem 1 for general covers. In general, problem (11) is an infinite-dimensional linear program. By assigning dual variables $f_r(\tilde{\boldsymbol{c}}_r)$, $d_{j,r}$ and $g_{j,r}(\tilde{\boldsymbol{c}}_{K_r})$ to the equalities in formulation (11), its dual problem is as follows:

$$
\begin{aligned}
M_{\mathcal{E}}^{\mathrm{D}}(\varphi_N) = \inf \quad & \sum_{r\in\mathcal{R}} \mathbb{E}_{\theta_r}\left[f_r(\tilde{\boldsymbol{c}}_r)\right] \\
\text{s.t.} \quad & f_r(\tilde{\boldsymbol{c}}_r) \geq \tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r \circ \boldsymbol{a}_{\boldsymbol{j}_r}) - g_{j,r}(\tilde{\boldsymbol{c}}_{K_r}) + \sum_{t>r:\sigma_t=r} g_{j,t}(\tilde{\boldsymbol{c}}_{K_t}) - d_{j,r}, \quad \forall j \in \mathcal{M}, \forall r \in \mathcal{R}, \forall \boldsymbol{c}_r \in B_r, \\
& \sum_{r\in\mathcal{R}} d_{j,r} + b_j = 0, \quad \forall j \in \mathcal{M},
\end{aligned} \tag{13}
$$

which can be concisely written as:

$$M_{\mathcal{E}}^{\mathrm{D}}(\varphi_N) = \inf_{g_{j,r}(\boldsymbol{c}_{K_r}),d_{jr}} \quad \sum_{r\in\mathcal{R}} \mathbb{E}_{\theta_r}\left[\max_{j\in\mathcal{M}}\left(\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r \circ \boldsymbol{a}_{\boldsymbol{j}_r}) - g_{j,r}(\tilde{\boldsymbol{c}}_{K_r}) + \sum_{t>r:\sigma_t=r} g_{j,t}(\tilde{\boldsymbol{c}}_{K_t}) - d_{j,r}\right)\right]$$
$$\text{s.t.} \quad \sum_{r\in\mathcal{R}} d_{j,r} + b_j = 0, \quad \forall j \in \mathcal{M}, \tag{14}$$

where we define $K_1 \triangleq \emptyset$, and for $r \in \mathcal{R}$, if $K_r = \emptyset$, we define $\tilde{\boldsymbol{c}}_{K_r} \triangleq 0$, and $g_{j,r}(0) \triangleq 0$, for all $j \in \mathcal{M}$. Weak duality implies that $M_{\mathcal{E}}^{\mathrm{D}}(\varphi_N)$ is an upper bound of $M_{\mathcal{E}}^{\mathrm{P}}(\varphi_N)$ or equivalently, $M_{\mathcal{E}}(\varphi_N)$. Since φ_N is a bounded continuous function in $\tilde{\boldsymbol{c}}$ over bounded domains, strong duality holds under mild conditions, for example, bounded supports B_r for $\tilde{\boldsymbol{c}}_r$, $r \in \mathcal{R}$ (see Rüschendorf [38], Kellerer [18] and references therein). Under these conditions, $M_{\mathcal{E}}(\varphi_N)$ is equivalent to the dual objective of $M_{\mathcal{E}}^{\mathrm{D}}(\varphi_N)$. For a general partition, the dual variables $g_{j,r}(\boldsymbol{c}_{K_r})$ in (13) correspond to the marginal consistency constraints in the primal problem (11). When $\mathcal{E}$ is a partition, the marginal consistency constraints are not needed and hence neither are the corresponding dual variables. Thus formulation (13) reduces to

$$M_{\mathcal{E}}^{\mathrm{D}}(\varphi_N) = \inf \quad \sum_{r\in\mathcal{R}} \mathbb{E}_{\theta_r}\left[\max_{j\in\mathcal{M}}\left(\boldsymbol{c}_r^T \boldsymbol{a}_{\boldsymbol{j}_r} - d_{j,r}\right)\right]$$
$$\text{s.t.} \quad \sum_{r\in\mathcal{R}} d_{j,r} + b_j = 0, \quad \forall j \in \mathcal{M}, \tag{15}$$

which is equivalent to the non-overlapping marginal formulation in Doan and Natarajan [4].

2.2 Connected Regular Covers: Star and Series Case

Star and series covers are examples of overlapping regular *connected* covers. A cover $\mathcal{E}$ is said to be connected if for any $s \neq t \in \mathcal{R}$, there exists a sequence $r_1 = s, r_2, \ldots, r_m = t \in \mathcal{R}$ such that $J_{r_j} \cap J_{r_{j+1}} \neq \emptyset$ for all $j = 1, ..., m-1$. It is clear that partitions are not connected covers. The next lemma characterizes the connectedness of regular covers.

Lemma 2 *A regular cover $\mathcal{E}$ is connected if and only if $K_r \neq \emptyset$ for all $r \in \mathcal{R} \setminus \{1\}$.*

The proof of the Lemma is provided in the Appendix. This characterization of connected regular covers allows us to simplify the formulations to compute $M_{\mathcal{E}}(\varphi_N)$ as shown in the following proposition.

Proposition 1 *Given a connected regular cover $\mathcal{E}$ and a consistent set $\{\theta_r\}_{r\in\mathcal{R}}$ of marginal distributions, define*

$$
\begin{aligned}
M_{\mathcal{E}}^{\mathrm{PC}}(\varphi_N) = \sup_{\vartheta_{j,r}(\boldsymbol{c}_r)} \quad & \sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{M}}\int_{B_r}\left(\boldsymbol{c}_r^T(\boldsymbol{\eta}_r\circ \boldsymbol{a}_{\boldsymbol{j}_r})+\rho_r b_j\right)\cdot\vartheta_{j,r}(\boldsymbol{c}_r)\ \mathrm{d}\boldsymbol{c}_r \\
\text{s.t.} \quad & \vartheta_{j,r}(\boldsymbol{c}_r)\geq 0, && \forall r\in\mathcal{R},\forall \boldsymbol{c}_r\in B_r,\forall j\in\mathcal{M}, \\
& \sum_{j\in\mathcal{M}}\vartheta_{j,r}(\boldsymbol{c}_r)=\theta_r(\boldsymbol{c}_r), && \forall r\in\mathcal{R},\forall \boldsymbol{c}_r\in B_r, \\
& proj_{K_r}(\vartheta_{j,r}(\boldsymbol{c}_{K_r}))=proj_{K_r}(\vartheta_{j,\sigma_r}(\boldsymbol{c}_{K_r})), && \forall r\in\mathcal{R}\setminus\{1\},\forall \boldsymbol{c}_{K_r}\in B_{K_r}, \\
& && \forall j\in\mathcal{M},
\end{aligned}
\tag{16}
$$

where ρ_r, $r\in\mathcal{R}$, are arbitrary constants that satisfy $\sum_{r\in\mathcal{R}}\rho_r=1$. Then $M_{\mathcal{E}}(\varphi_N)=M_{\mathcal{E}}^{\mathrm{PC}}(\varphi_N)$.

Proof. We claim that the constraints

$$\int_{B_r}\vartheta_{j,r}(\boldsymbol{c}_r)\ \mathrm{d}\boldsymbol{c}_r=\lambda_j,\quad \forall r\in\mathcal{R},\forall j\in\mathcal{M},$$

in (11) are redundant if the regular cover $\mathcal{E}$ is connected. From Lemma 2, $K_r\neq\emptyset$ for all $r\in\mathcal{R}$. Using the last set of constraints in (11) (and (16)), we obtain the following equalities:

$$\int_{B_r}\vartheta_{j,r}(\boldsymbol{c}_r)\ \mathrm{d}\boldsymbol{c}_r=\int_{B_s}\vartheta_{j,s}(\boldsymbol{c}_s)\ \mathrm{d}\boldsymbol{c}_s,\quad \forall r\neq s\in\mathcal{R},\forall j\in\mathcal{M},$$

which shows that this constraint is redundant. Thus, we can drop the decision variables λ_j, $j\in\mathcal{M}$, by replacing $\sum_{j\in\mathcal{M}}b_j\lambda_j$ by $\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{M}}\int_{B_r}\rho_r b_j\vartheta_{j,r}(\boldsymbol{c}_r)\mathrm{d}\boldsymbol{c}_r$ in the objective given that $\sum_{r\in\mathcal{R}}\rho_r=1$. Thus for connected regular covers, (16) is equivalent to (11) and it implies that in this case, $M_{\mathcal{E}}(\varphi_N)=M_{\mathcal{E}}^{\mathrm{PC}}(\varphi_N)$. □

The problem (16) has fewer variables and lesser constraints as compared to (11), which also allows us to simplify its dual formulation by remove the corresponding set of dual variables. The dual formulation in this case is written as follows:

$$M_{\mathcal{E}}^{\mathrm{DC}}(\varphi_N)=\inf_{g_{j,r}(\boldsymbol{c}_{K_r})}\sum_{r\in\mathcal{R}}\mathbb{E}_{\theta_r}\left[\max_{j\in\mathcal{M}}\left(\tilde{\boldsymbol{c}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{a}_{\boldsymbol{j}_r})-g_{j,r}\left(\tilde{\boldsymbol{c}}_{K_r}\right)+\sum_{t>r:\sigma_t=r}g_{j,t}\left(\tilde{\boldsymbol{c}}_{K_t}\right)+\rho_r b_j\right)\right]. \tag{17}$$

We consider two simple examples of the Fréchet bound for star and series covers.

Series cover: If $\mathcal{E} = \mathcal{E}^+$ is a series cover, we have $R = N-1$ and $\sigma_r = r-1, \forall r = 2, ..., N-1$. Letting $\rho_r = 1/(N-1)$ for all $r \in \mathcal{R}$, we obtain the following formulation directly from (17):

$$\begin{aligned} M_{\mathcal{E}^+}^{\mathrm{DC}}(\varphi_N) = \inf_{g_{j,r}(\tilde{c}_r)} \quad & \mathbb{E}_{\theta_{1,2}}\left[\max_{j\in\mathcal{M}}\left(\tilde{c}_1 a_{j_1} + \frac{\tilde{c}_2 a_{j_2}}{2} + g_{j,2}(\tilde{c}_2) + \frac{b_j}{N-1}\right)\right] \\ & + \sum_{r=2}^{N-2} \mathbb{E}_{\theta_{r,r+1}}\left[\max_{j\in\mathcal{M}}\left(\frac{\tilde{c}_r a_{j_r}}{2} + \frac{\tilde{c}_{r+1} a_{j_{r+1}}}{2} - g_{j,r}(\tilde{c}_r) + g_{j,r+1}(\tilde{c}_{r+1}) + \frac{b_j}{N-1}\right)\right] \\ & + \mathbb{E}_{\theta_{N-1,N}}\left[\max_{j\in\mathcal{M}}\left(\tilde{c}_N a_{j_N} + \frac{\tilde{c}_{N-1} a_{j_{N-1}}}{2} - g_{j,N-1}(\tilde{c}_{N-1}) + \frac{b_j}{N-1}\right)\right]. \end{aligned} \tag{18}$$

Star cover: If $\mathcal{E} = \mathcal{E}^\star$ is a star cover, we have $R = N-1$ and $\sigma_r = 1, \forall r = 2, ..., N-1$. We set $\eta_i^1 = 1$ and $\eta_i^r = 0$ for all $i \in I_0$ and $r \in \mathcal{R} \setminus \{1\}$. Similarly, we set $\rho_1 = 1$ and $\rho_r = 0$ for all $r \in \mathcal{R} \setminus \{1\}$. The dual Fréchet bound for the star cover $\mathcal{E}^\star$ can then be computed as follows:

$$\begin{aligned} M_{\mathcal{E}^\star}^{\mathrm{DC}}(\varphi_N) = \inf \quad & \mathbb{E}_{\theta_{1,0}}\left[\max_{j\in\mathcal{M}}\left(\tilde{c}_1 a_{j1} + \tilde{c}_0 a_{j_0} + \sum_{r=2}^{N-1} g_{j,r}(\tilde{c}_0) + b_j\right)\right] \\ & + \sum_{r=2}^{N-1} \mathbb{E}_{\theta_{r,0}}\left[\max_{j\in\mathcal{M}}\left(\tilde{c}_r a_{j_r} - g_{j,r}(\tilde{c}_0)\right)\right]. \end{aligned}$$

We introduce the following new set of decision variables:

$$g_{j,1}(c_0) \triangleq -\sum_{r=2}^{N-1} g_{j,r}(c_0) - c_0 a_{j_0} - b_j, \quad \forall c_0 \in B_0, \forall j \in \mathcal{M}.$$

The formulation then becomes:

$$\begin{aligned} M_{\mathcal{E}^\star}^{\mathrm{DC}}(\varphi_N) = \inf \quad & \sum_{r\in\mathcal{R}} \mathbb{E}_{\theta_{r,0}}\left[\max_{j\in\mathcal{M}}\left(\tilde{c}_r a_{j_r} - g_{j,r}(\tilde{c}_0)\right)\right] \\ \text{s.t.} \quad & \sum_{r\in\mathcal{R}} g_{j,r}(c_0) = -c_0 a_{j_0} - b_j, \quad \forall c_0 \in B_0, \forall j \in \mathcal{M}. \end{aligned} \tag{19}$$

Using a conditioning trick (see Puccetti and Ruschendorf [31]), we can express (19) as follows:

$$M_{\mathcal{E}^\star}^{\mathrm{DC}}(\varphi_N) = \mathbb{E}_{\theta_0}\left\{\begin{array}{ll} \inf & \sum_{r\in\mathcal{R}} \mathbb{E}_{\theta_r}\left[\max_{j\in\mathcal{M}}\left(\tilde{c}_r a_{j_r} - g_{j,r}(\tilde{c}_0)\right)\right] \\ \text{s.t.} & \sum_{r\in\mathcal{R}} g_{j,r}(\tilde{c}_0) = -\tilde{c}_0^T a_{j_0} - b_j, \quad \forall j \in \mathcal{M}. \end{array}\right\},$$

where θ_r denotes the marginal distribution of $\tilde{\boldsymbol{c}}_r$ that can be computed from $\theta_{r,0}$ for all $r \in \mathcal{R}$. For a fixed $c_0 \in B_0$, the inner problem is the Fréchet bound of

$$\mathbb{E}_\mu\left[\max_{i=1,\ldots,M}\left(\sum_{r\in\mathcal{R}} \tilde{c}_r a_{j_r} + \left(c_0^T a_{j_0} + b_j\right)\right)\right],$$

where μ belongs to the Fréchet class $\Theta_{\mathcal{P}(\theta_1,\ldots,\theta_r)}$ defined by the partition $\mathcal{P}$ of $(\tilde{\boldsymbol{c}}_1,\ldots,\tilde{\boldsymbol{c}}_R)$. Thus, the Fréchet bound for the star cover $\mathcal{E}^\star$ can be written as follows:

$$M_{\mathcal{E}^\star}^{\mathrm{DC}}(\varphi_N) = \mathbb{E}_{\theta_0}\left\{ \sup_{\mu\in\Theta_{\mathcal{P}(\theta_1,\ldots,\theta_r)}} \mathbb{E}_\mu\left[\varphi_N(\tilde{\boldsymbol{c}})\right]\right\}, \tag{20}$$

which indicates that its computation can be reduced to the computation of Fréchet bounds with general partitions.

3 Bounds for CVaR and VaR

In this section, we derive new Fréchet bounds of CVaR and VaR from the results of the previous section.

3.1 CVaR Bound on Sum of Random Variables

The distributionally robust bound on CVaR is the worst-case CVaR with respect to the corresponding Fréchet class:

$$\mathrm{WCVaR}_\alpha(\boldsymbol{x}) \triangleq \sup_{\theta\in\Theta_{\mathcal{E}}} \mathrm{CVaR}_\alpha^\theta(\boldsymbol{x}).$$

where

$$\mathrm{CVaR}_\alpha^\theta(\boldsymbol{x}) = \inf_{\beta\in\Re}\left\{\beta + \frac{1}{1-\alpha}\mathbb{E}_\theta\left[\left(\sum_{i\in\mathcal{N}}\tilde{c}_i x_i - \beta\right)^+\right]\right\}, \quad \alpha\in(0,1).$$

Worst-case CVaR has been studied for different classes of distributions (see, e.g., Natarajan, Sim and Uichanco [29] and Zhu and Fukushima [45]). Under the assumption that all the distributions $\{\theta_r\}_{r\in\mathcal{R}}$ have finite second moments, we can interchange the infimum and supremum in the worst-case CVaR formulation due to the finiteness of the expected value and the convexity of the objective function with respect to β and linearity with respect to the measure θ:

$$\begin{aligned}\mathrm{WCVaR}_\alpha(\boldsymbol{x}) &= \sup_{\theta\in\Theta_{\mathcal{E}}}\inf_{\beta\in\Re}\left\{\beta + \frac{1}{1-\alpha}\mathbb{E}_\theta\left[\left(\sum_{i\in\mathcal{N}}\tilde{c}_i x_i - \beta\right)^+\right]\right\} \\ &= \inf_{\beta\in\Re}\left\{\beta + \frac{1}{1-\alpha}\sup_{\theta\in\Theta_{\mathcal{E}}}\mathbb{E}_\theta\left[\left(\sum_{i\in\mathcal{N}}\tilde{c}_i x_i - \beta\right)^+\right]\right\}.\end{aligned} \tag{21}$$

Thus, in order to compute the Fréchet bound of CVaR, we need to compute an upper bound of the expected value of the function $\varphi_t(\boldsymbol{c};\boldsymbol{x}) = \left(\sum_{i\in\mathcal{N}}\tilde{c}_i x_i - \beta\right)^+$. Consider the case of a sum of random variables with $\boldsymbol{x} = \boldsymbol{e}$, the vector of all ones. The inner problem is written as follows:

$$M_{\mathcal{E}}(\varphi_t) = \sup_{\theta \in \Theta_{\mathcal{E}}} \mathbb{E}_\theta \left[\left(\sum_{i \in \mathcal{N}} \tilde{c}_i - \beta \right)^+ \right].$$

Letting $M = 2$, $\boldsymbol{a}_1 = \boldsymbol{e}$, $b_1 = -\beta$, and $\boldsymbol{a}_2 = \mathbf{0}$, $b_2 = 0$, we can directly obtain the following formulation from (13):

$$\begin{aligned} M_{\mathcal{E}}^{\mathrm{D}}(\varphi_t) = \inf_{g_r(\boldsymbol{c}_{K_r}), d_r} \quad & \sum_{r \in \mathcal{R}} \mathbb{E}_{\theta_r} \left[\left(\tilde{\boldsymbol{c}}_r^T \boldsymbol{\eta}_r - g_r(\tilde{\boldsymbol{c}}_{K_r}) + \sum_{t > r : \sigma_t = r} g_t(\tilde{\boldsymbol{c}}_{K_t}) - d_r \right)^+ \right] \\ \text{s.t.} \quad & \sum_{r \in \mathcal{R}} d_r = \beta. \end{aligned} \tag{22}$$

Then the worst-case CVaR is computed as follows:

$$\begin{aligned} \mathrm{WCVaR}_\alpha(\boldsymbol{e}) = \inf_{g_r(\boldsymbol{c}_{K_r}), d_r} \quad & \beta + \frac{1}{1-\alpha} \sum_{r \in \mathcal{R}} \mathbb{E}_{\theta_r} \left[\left(\tilde{\boldsymbol{c}}_r^T \boldsymbol{\eta}_r - g_r(\tilde{\boldsymbol{c}}_{K_r}) + \sum_{t > r : \sigma_t = r} g_t(\tilde{\boldsymbol{c}}_{K_t}) - d_r \right)^+ \right] \\ \text{s.t.} \quad & \sum_{r \in \mathcal{R}} d_r = \beta. \end{aligned} \tag{23}$$

3.1.1 Example

Our next example provides a comparison of the maximum entropy distribution in (8) and the worst-case distribution for CVaR. In this experiment, we compare the two expected values:

$$\sup_{\theta \in \Theta_{\mathcal{E}}} \mathbb{E}_\theta \left[\left(\sum_{i \in \mathcal{N}} \tilde{c}_i - \beta \right)^+ \right] \longleftrightarrow \mathbb{E}_{\mathrm{ME}} \left[\left(\sum_{i \in \mathcal{N}} \tilde{c}_i - \beta \right)^+ \right],$$

where ME is the maximum-entropy distribution in (8). Let $B = \{0,1\}^4$ be the support of $\tilde{\boldsymbol{c}}$ and the cover $\mathcal{E}$ be the simple series cover. Consider the setting with uniform discrete marginal distributions, i.e.,

$$P_{\{1,2\}}(\boldsymbol{c}) = P_{\{2,3\}}(\boldsymbol{c}) = P_{\{3,4\}}(\boldsymbol{c}) = 1/4, \quad \forall \boldsymbol{c} \in \{0,1\}^2.$$

In this setting, the random variables in each subset are actually independent. The maximum-entropy distribution is easy to compute from the formulation in (8) as the independent uniform distribution:

$$P_{\mathrm{ME}}(\boldsymbol{c}) = 1/16, \quad \forall \boldsymbol{c} \in \{0,1\}^4.$$

The worst-case formulation (22) is a linear optimization problem that can be solved efficiently. Since $B = \{0,1\}^4$, the support of $\sum_{i \in \mathcal{N}} \tilde{c}_i$ is $\{0,1,2,3,4\}$ and hence we restrict our attention to $\beta \in [0,4]$. The computational results are shown in Figure 1. The Fréchet bound provides an upper bound on the expected value computed with respect to the maximum-entropy distribution. Note that to compute the expected value for the maximum entropy distribution might need an exponential number of arithmetic operations for arbitrary supports while the worst-case bound is computable in polynomial time.

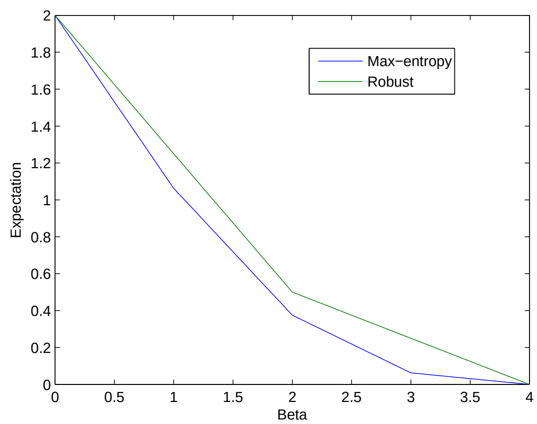

Figure 1: Comparison between the Fréchet bound the maximum entropy distribution

3.2 VaR bound on sum of random variables

For the sum of random variable, the value-at-risk (VaR) is defined as follows:

$$\mathrm{VaR}_\alpha^{\theta}(\boldsymbol{x}) \triangleq \inf \left\{ z \in \Re : \mathbb{P}_\theta \left(\sum_{i \in \mathcal{N}} \tilde{c}_i x_i \le z \right) \ge \alpha \right\}, \quad \alpha \in (0,1).$$

Then, we have the following equivalence:

$$\mathrm{VaR}_\alpha^{\theta}(\boldsymbol{x}) \le z \;\Leftrightarrow\; \mathbb{P}_\theta \left(\sum_{i \in \mathcal{N}} \tilde{c}_i x_i \le z \right) \ge \alpha,$$

which implies that z is an upper bound of $\mathrm{VaR}_\alpha^{\theta}(\boldsymbol{x})$, if and only if α is a lower bound of the cumulative distribution function value of the sum risk, $\mathbb{P}_\theta \left(\sum_{i \in \mathcal{N}} \tilde{c}_i x_i \le z \right)$. Since CVaR dominates

VaR, i.e., $\text{VaR}_\alpha^\theta(\boldsymbol{x}) \leq \text{CVaR}_\alpha^\theta(\boldsymbol{x})$ (see Rockafellar and Uryasev [34]), we can use CVaR to derive lower bounds of the cumulative distribution function values.

Given a Fréchet class $\Theta_\mathcal{E}$ of distributions, the worst-case VaR is defined as follows:

$$\text{WVaR}_\alpha(\boldsymbol{x}) \quad \triangleq \quad \inf\left\{ z \in \Re : \inf_{\theta\in\Theta_\mathcal{E}} \mathbb{P}_\theta\left(\sum_{i\in\mathcal{N}} \tilde{c}_i x_i \leq z\right) \geq \alpha \right\}. \tag{24}$$

Since $\text{WVaR}_\alpha(\boldsymbol{x}) = \sup_{\theta\in\Theta_\mathcal{E}} \text{WVaR}_\alpha^\theta(\boldsymbol{x})$, this implies $\text{WVaR}_\alpha(\boldsymbol{x}) \leq \text{WCVaR}_\alpha(\boldsymbol{x})$. Hence,

$$\text{WVaR}_\alpha(\boldsymbol{e}) \leq z \;\Leftrightarrow\; \inf_{\theta\in\Theta_\mathcal{E}} \mathbb{P}_\theta\left(\sum_{i\in\mathcal{N}} \tilde{c}_i \leq z\right) = m_\mathcal{E}(\varphi_d) \geq \alpha,$$

where $\varphi_d(\tilde{\boldsymbol{c}}; x) = \mathbb{I}\left\{\sum_{i\in\mathcal{N}} \tilde{c}_i \leq x\right\}$. It is clear that φ_d is related to φ_d^- and so are their Fréchet bounds, $m_\mathcal{E}(\varphi_d^-(.;x)) \leq m_\mathcal{E}(\varphi_d(.;x)) \leq m_\mathcal{E}(\varphi_d^-(.;x+\epsilon))$ for all $\epsilon > 0$. Since $\text{WCVaR}_\alpha(\boldsymbol{x})$ is an upper bound of $\text{WVaR}_\alpha(\boldsymbol{x})$, we compute it using (23) to find lower bounds of $m_\mathcal{E}(\varphi_d)$ and compare them with standard bound of $m_\mathcal{E}(\varphi_d^-)$ in Embrechts and Puccetti [8].

3.2.1 Example

In the following numerical experiment, we consider the simple series cover. Using (23), we can compute the worst-case CVaR as follows:

$$\begin{aligned}
\text{WCVaR}_\alpha(\boldsymbol{e}) = \inf\nolimits_{g_i(c_i)} \quad & \beta + \frac{1}{1-\alpha}\left\{ \mathbb{E}_{\theta_{1,2}}\left[\left(\tilde{c}_1 + \frac{\tilde{c}_2}{2} + g_2(\tilde{c}_2) - \frac{\beta}{N-1}\right)^+\right] \right. \\
& + \sum_{i=2}^{N-2} \mathbb{E}_{P_{i,i+1}}\left[\left(\frac{\tilde{c}_i + \tilde{c}_{i+1}}{2} - g_i(\tilde{c}_i) + g_{i+1}(\tilde{c}_{i+1}) - \frac{\beta}{N-1}\right)^+\right] \\
& \left. + \mathbb{E}_{P_{N-1,N}}\left[\left(\frac{\tilde{c}_{N-1}}{2} + \tilde{c}_N - g_{N-1}(\tilde{c}_{N-1}) - \frac{\beta}{N-1}\right)^+\right]\right\}.
\end{aligned} \tag{25}$$

We construct bivariate marginal distributions for the simple series cover by using the independent copula and identical univariate marginals, which are uniform distributions in $[0,1]$. Then $F_i(c_i) = F(c_i)$ for all $i = 1,\ldots,N$, where $F(c_i) = c_i$ for all $c_i \in [0,1]$ and the joint distribution for two random variables is is $F_{i,i+1}(c_i, c_{i+1}) = c_i c_{i+1}$ for all $(c_i, c_{i+1}) \in [0,1]^2$, $i = 1,\ldots,N-1$. It is clear that this set of these bivariate marginals is consistent. Given these continuous marginals, the problem in (25) is an infinite-dimensional linear optimization problem. To compute $\text{WCVaR}_\alpha(\boldsymbol{e})$, we use a discretization approach for the cumulative distribution function F to compute upper and lower bounds. Consider a discrete distribution $\hat{F}_\omega$ approximation of F with M-points as in Embrechts and Puccetti [8]:

$$\hat{F}_{\boldsymbol{\omega}} \triangleq \frac{1}{M} \sum_{j \in \mathcal{M}} \mathbb{I}\{x \geq \omega_j\},$$

where $\boldsymbol{\omega} = \{\omega_1, ..., \omega_M\}$ is the set of M jump points. Letting $q_j = \dfrac{j}{M}$, $j = 0, \ldots, M$, define two sets of jump points $\overline{\boldsymbol{\omega}} = \{q_1, ..., q_M\}$ and $\underline{\boldsymbol{\omega}} = \{q_0, ..., q_{M-1}\}$. Clearly, $\hat{F}_{\overline{\boldsymbol{\omega}}}$ and $\hat{F}_{\underline{\boldsymbol{\omega}}}$ provide lower and upper bounds for F. The discretized bivariate marginal distributions $P_{\overline{\boldsymbol{\omega}}}$ and $P_{\underline{\boldsymbol{\omega}}}$ are simply constructed from the corresponding discretized univariate marginals using the independent copula. Let $\text{WCVaR}_\alpha(\boldsymbol{e})$ denote the worst-case CVaR with respect to the original continuous marginals and $\overline{\text{WCVaR}}_\alpha(\boldsymbol{e})$ and $\underline{\text{WCVaR}}_\alpha(\boldsymbol{e})$ denote the bounds with respect to the discretized marginals $P_{\overline{\boldsymbol{\omega}}}$ and $P_{\underline{\boldsymbol{\omega}}}$, respectively. We then have $\underline{\text{WCVaR}}_\alpha(\boldsymbol{e}) \leq \text{WCVaR}_\alpha(\boldsymbol{e}) \leq \overline{\text{WCVaR}}_\alpha(\boldsymbol{e})$. The upper and lower bounds $\overline{\text{WCVaR}}_\alpha(\boldsymbol{e})$ and $\underline{\text{WCVaR}}_\alpha(\boldsymbol{e})$ can be computed as optimal values of linear optimization problems derived from (25).

Embrechts and Puccetti [8] computed a lower bound of $m_{\mathcal{E}}(\varphi_d^-)$ using the standard bound in (6) with variable splitting. In order to compute this bound, one needs to calculate $F_y(d) = \mathbb{P}\left(\tilde{c}_1 + \dfrac{\tilde{c}_2}{2} \leq d\right)$ and $F_z(d) = \mathbb{P}\left(\dfrac{\tilde{c}_1}{2} + \dfrac{\tilde{c}_2}{2} \leq d\right)$. In our example, this reduces to:

$$F_y(d) = \begin{cases} 0, & d < 0, \\ d^2, & 0 \leq d < 1/2, \\ d - 1/4, & 1/2 \leq d < 1, \\ -d^2 + 3d - 5/4, & 1 \leq d < 3/2, \\ 1, & d \geq 3/2, \end{cases} \quad \text{and} \quad F_z(d) = \begin{cases} 0, & d < 0, \\ 2d^2, & 0 \leq d < 1/2, \\ -2d^2 + 4d - 1, & 1/2 \leq d < 1, \\ 1, & d \geq 1. \end{cases}$$

Note that $F_y^-(d) = F_y(d)$ and $F_z^-(d) = F_z(d)$ given continuous distributions. Thus the lower bound of $m_{\mathcal{E}}(\varphi_d^-(.; x))$ from Embrechts and Puccetti [8] can be computed as follows:

$$\text{RSB}(x) = \max\left\{ \sup_{\boldsymbol{d} \in \Re^{N-2}} \left[F_y(d_1) + F_y\left(x - \sum_{i=1}^{N-2} d_i\right) + \sum_{i=2}^{N-2} F_z(d_i) \right] - (N-2), 0 \right\}. \tag{26}$$

The inner problem is a maximization problem, with the objective function

$$H(\boldsymbol{d}; x) = F_y(d_1) + F_y\left(x - \sum_{i=1}^{N-2} d_i\right) + \sum_{i=2}^{N-2} F_z(d_i) - (N-2),$$

which is unfortunately not a concave function. In general, it is hard to find optimal solutions for a non-convex optimization problem. For this particular problem, we solve it numerically using the procedure outlined in the Appendix. We use $M = 50$ to numerically compute the two bounds $\overline{\mathrm{WCVaR}}_\alpha(\boldsymbol{e})$ and $\underline{\mathrm{WCVaR}}_\alpha(\boldsymbol{e})$ of $\mathrm{WCVaR}_\alpha(\boldsymbol{e})$. We then compute the two bounds $\mathrm{SECB}^+(x)$ and $\mathrm{SECB}^-(x)$ for $m_{\mathcal{E}}(\varphi_d)$, which are the inverse of $\overline{\mathrm{WCVaR}}_\alpha(\boldsymbol{e})$ and $\underline{\mathrm{WCVaR}}_\alpha(\boldsymbol{e})$, respectively.

Figure 2 shows three bounds, $\mathrm{RSB}(x)$, $\mathrm{SECB}^+(x)$, and $\mathrm{SECB}^-(x)$, for $N = 4$ and $N = 6$. We can see that $\mathrm{SECB}^+(x)$ and $\mathrm{SECB}^-(x)$ are close to each other. Since the actual CVaR bound $\mathrm{SE}(\alpha)$ lies in the interval $\left[\underline{\mathrm{SE}}(\alpha); \overline{\mathrm{SE}}(\alpha)\right]$, we can claim that the discrete approximation with $M = 50$ for CVaR bounds is reasonably good in this example. It is also clear that our proposed approximation significantly improves on the existing standard bound.

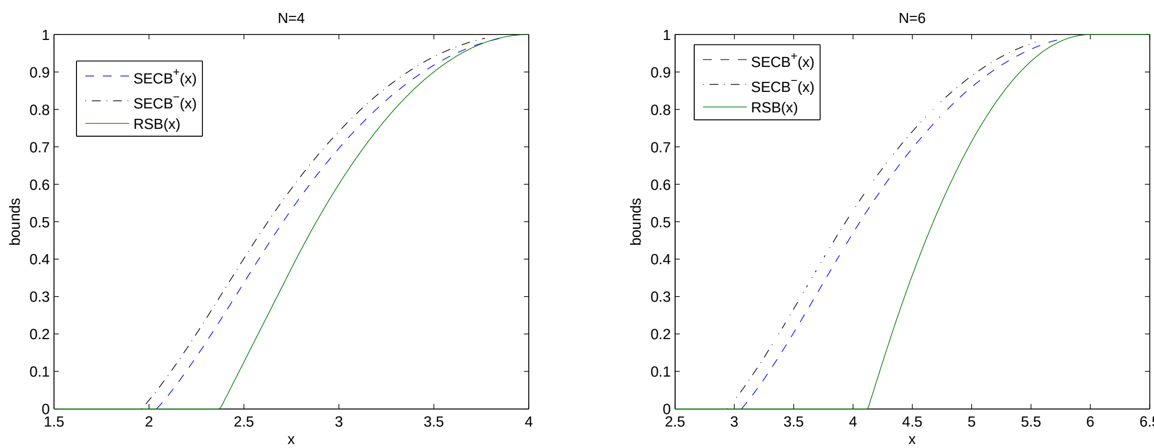

Figure 2: Different bounds of $m_{\mathcal{E}}(\varphi_d)$ with the simple series cover

4 Robust Portfolio Optimization

In this section, we implement a distributionally robust portfolio optimization problem as introduced in Section 1. Consider a portfolio of N assets and let $\tilde{\boldsymbol{\xi}}$ be the random return vector of all assets. The random loss of ith asset is then simply $\tilde{c}_i = -\tilde{\xi}_i$, $i \in \mathcal{N}$. Given a feasible asset allocation $\boldsymbol{x} \in \mathcal{X}$, the computation of CVaR of the joint portfolio loss requires the distribution of the random return vector $\tilde{\boldsymbol{\xi}}$. In the financial market, we have access to historical data of say S samples. The sample distribution θ is defined using the historical data as $\left\{\tilde{\boldsymbol{\xi}}_1, \ldots, \tilde{\boldsymbol{\xi}}_S\right\}$ where each sample vector occurs with probability $1/S$. The optimal sample based allocation with the minimum CVaR is obtained by solving the following optimization problem:

$$\inf_{\beta\in\Re,\boldsymbol{x}\in\mathcal{X}} \left\{\beta + \frac{1}{(1-\alpha)S}\sum_{s=1}^{S}\left[\left(-\tilde{\boldsymbol{\xi}}_s^T\boldsymbol{x}-\beta\right)^+\right]\right\}. \tag{27}$$

However, the out-of-sample performance of such an approach is not necessarily good due to the possibility of over-fitting where the out-of-sample distribution is different from the in-sample distribution. Lim, Shanthikumar and Vahn [1] show using simulated data that the CVaR measure with sample based optimization results in fragile portfolios that are often unreliable due to estimation errors. In the distributional robust optimization approach, we use the historical data to extract the stable dependencies among the random losses and incorporate only this information into the optimization model. Given the historical data of asset returns, we construct a Fréchet class of distributions $\Theta_{\mathcal{E}}$ and consider the following robust optimization problem:

$$\inf_{\boldsymbol{x}\in\mathcal{X}}\sup_{\theta\in\Theta_{\mathcal{E}}} \mathrm{CVaR}_{\alpha}^{\theta}(\boldsymbol{x}). \tag{28}$$

Using the dual representation in (23), the distributional robust portfolio optimization problem is formulated as:

$$\begin{aligned}\inf_{d_r,g_r(\boldsymbol{\xi}_{K_r})}\quad & \beta+\frac{1}{1-\alpha}\sum_{r\in\mathcal{R}}\mathbb{E}_{\theta_r}\left[\left(-\tilde{\boldsymbol{\xi}}_r^T(\boldsymbol{\eta}_r\circ\boldsymbol{x}_r)-g_r\left(\tilde{\boldsymbol{\xi}}_{K_r}\right)+\sum_{t>r:\sigma_t=r}g_t\left(\tilde{\boldsymbol{\xi}}_{K_t}\right)-d_r\right)^+\right]\\ \text{s.t.}\quad & \sum_{r\in\mathcal{R}}d_r=\beta,\\ & \boldsymbol{x}\in\mathcal{X}.\end{aligned} \tag{29}$$

If $\mathcal{X}$ is a polyhedron and the multivariate marginals are discrete distributions, problem (29) is a linear optimization problem which can be solved efficiently. We discuss a data-driven approach to construct the Fréchet class of distributions of asset returns $\Theta_{\mathcal{E}}$ next.

4.1 Construction of Overlapping Covers

In the context of distributionally robust optimization, the dependency structure of the random parameters is typically incorporated using moment information. Common classes of distributions used in the literature are specified with information on the first and second moments (see for example, El Ghaoui et al. [12] and Bertsimas et al. [2]) or multivariate normal distributions with parameter uncertainty in the mean and covariance matrix (see Garlappi, Uppal and Wang [21]). The resulting formulations are then tractable conic programs. Our approach is to use a Fréchet class of distributions with possibly overlapping marginals to capture partial information of dependencies among the random parameters. An important feature is to identify the cover structure $\mathcal{E}$ to balance over-fitting and getting overly conservative solutions due to the lack of information. In order to construct the cover $\mathcal{E}$, we use time-dependent correlation information of the asset returns. Under the assumption that the dependency structure that is stable is the one with minimum change in correlation over time, we propose the following data-driven approach to identify the Fréchet class of distributions:

Step 1: Split the historical data into two sets of equal size and compute the correlation matrices for the two data sets. By comparing correlation coefficients across the two matrices, we identify the pairs of assets with minimal differences in the correlation information of the returns. These pairs of returns are included in the same subset of the cover.

Step 2: Construct an undirected graph $G(V, E)$, where V is the set of all assets, and E is the set of all selected pairs of assets. If the graph $G(E, V)$ is chordal, i.e., each of its cycles of four or more vertices has at least a chord (an edge connecting two non-adjacent vertices in the cycle), then one can construct a regular cover $\mathcal{E}$ efficiently (see Lauritzen, Speed and Vijjayan [23]). A linear time lexicographic breadth-first search (L-BFS) algorithm is used to determine whether a graph is chordal and to construct the regular cover $\mathcal{E}$ (see Rose, Tarjan and Lueker [35]). If the graph $G(V, E)$ is not chordal, one adds in a set of additional edges, which are called the *fill-in* edges, to make the graph chordal. Even though the problem of finding the fill-in with the minimum number of edges is NP-complete (Yannakakis [44]), there are efficient algorithms to find fill-ins with reasonably small number of edges (see for example, Huang and Darwiche [13] and Natanzon, Shamir and Sharan [28]).

Given the cover $\mathcal{E}$, we consider consistent marginals for the subsets of the cover. While one could use the sample distributions from historical data for the marginals, in our data set this is not sufficiently large to get a non-trivial Fréchet class of distributions $\Theta_{\mathcal{E}}$. To tackle this situation, we round the samples in constructing the marginals. The approach we adopt is to cluster the historical data of each asset return into several clusters and to replace the sample data with the respective cluster mean. The marginals with the rounded samples are used in our optimization approach. Under this construction, the mean of rounded samples remains the same as that of original ones. However, the size of supports of marginal distributions is reduced and this reduces the computational time of (29). In the next section, we investigate the effects of this rounding procedure as well as the effects of using dependence structures in a numerical experiment with real financial market data.

4.2 Numerical Example

In this section, we compare the performance of different trading strategies. The data set analyzed consists of historical daily returns for an industry portfolio obtained from the Fama & French data library ([11]). The portfolio consists of NYSE, AMEX and NASDAQ stocks classified by industry. This include industries such as finance, health, textiles, food and machinery. A total of 4400 observations of daily return data was available in a period spanning approximately 15 years before the financial crisis, from August 18, 1989 to February 1, 2007. Consider an investor who plans to invest in a portfolio of these $N = 49$ risky assets. He would like to minimize the risk of his investment, while guaranteeing a certain level of average return by choosing an appropriate trading strategy. The investor re-balances his portfolio every 200 days. The trading strategy provides an easily implementable data-driven approach (see Natarajan et al. [29] for a similar experiment).

We divide the 4400 samples into 22 periods, with each period consisting of 200 days. The investor starts his investment from the beginning of the third period. From then on, at the beginning of each period, the investor uses the portfolio return data of the last two periods to make the decision on the portfolio allocation for the current period. We allow for short selling and consider the following set of allowable allocations:

$$\mathcal{X} = \left\{ \boldsymbol{x} \in \Re^n : \boldsymbol{e}^T \boldsymbol{x} = 1, \boldsymbol{\mu}^T \boldsymbol{x} \geq \mu_t \right\},$$

where μ_t is the target return and $\boldsymbol{\mu}$ is the expected return vector of all assets. We consider four different strategies to find the portfolio allocation $\boldsymbol{x}$:

1. **Sample-based approach (SB):** The sample distribution is used and the allocation $\boldsymbol{x}$ is computed by solving the problem (27).

2. **Rounded-sampled-based approach (RSB):** The rounded sample distribution is used in (27) instead of the original sample distribution. This strategy serves as a control group to verify the effects of the rounding procedure.

3. **Cover based approach:** Let $\Delta\rho_{(i,j)}$ be the absolute value of the difference in the correlation coefficients of two assets i and j from the rounded samples in the last two periods. Use this value as the weight for the edge (i, j) in a complete graph K_N of N vertices. In order to construct the graph $G(V, E)$, we consider two different approaches:

   (a) **Minimum spanning tree approach (MST):** Find a minimum spanning tree of this K_N graph and construct a cover $\mathcal{E}$ with $N-1$ two-element subsets corresponding to edges of the tree. It is clear that the cover $\mathcal{E}$ is regular. In special instances, this tree cover reduces to the simple star and based series covers.

   (b) **Edge budgeting based approach (EB):** To construct the graph $G(V, E)$, we remove all the edges except for a fraction $r_a \in [0, 1]$ of the total number of edges $N(N-1)/2$ edges that are chosen with smallest weights. If $r_a = 0$, EB reduces to univariate marginal case. If $r_a = 1$, it reduces to the RSB approach. We then use the L-BFS algorithm to test if the graph is chordal. If yes, the L-BFS algorithm also outputs the corresponding regular cover $\mathcal{E}$. Else, we use a fill-in algorithm to make the graph chordal and use L-BFS again to obtain the cover. In our numerical example, we use the MINWEIGHTELIMORDER function, which is based on the fill-in algorithm developed by Huang and Darwiche [13], from PMTK3, a Matlab toolkit for probabilistic modeling ([5]).

   The allocation $\boldsymbol{x}$ for the cover based approaches are computed by solving the problem (29).

In the experiments, we cluster the return data into 10 clusters. We use the R package CKMEANS.1D.1P, which is based on a k-means clustering dynamic programming algorithm in one dimension (see Wang et al. [43]). The target return μ_t is varied between 0.04% and 0.08%. For each target return, we apply the four trading strategies for 20 periods. We then compute the aggregate out-of-sample mean and out-of-sample CVaR. The numerical tests were conducted in 64-bit Matlab 2011a with CVX, a package for solving convex programs ([14]).

The top graph in Figure 3 shows the out-of-sample efficient frontiers of the EB trading strategy with different values of the parameter r_a when $\alpha = 0.9$. If $r_a = 0$, the EB strategy uses only

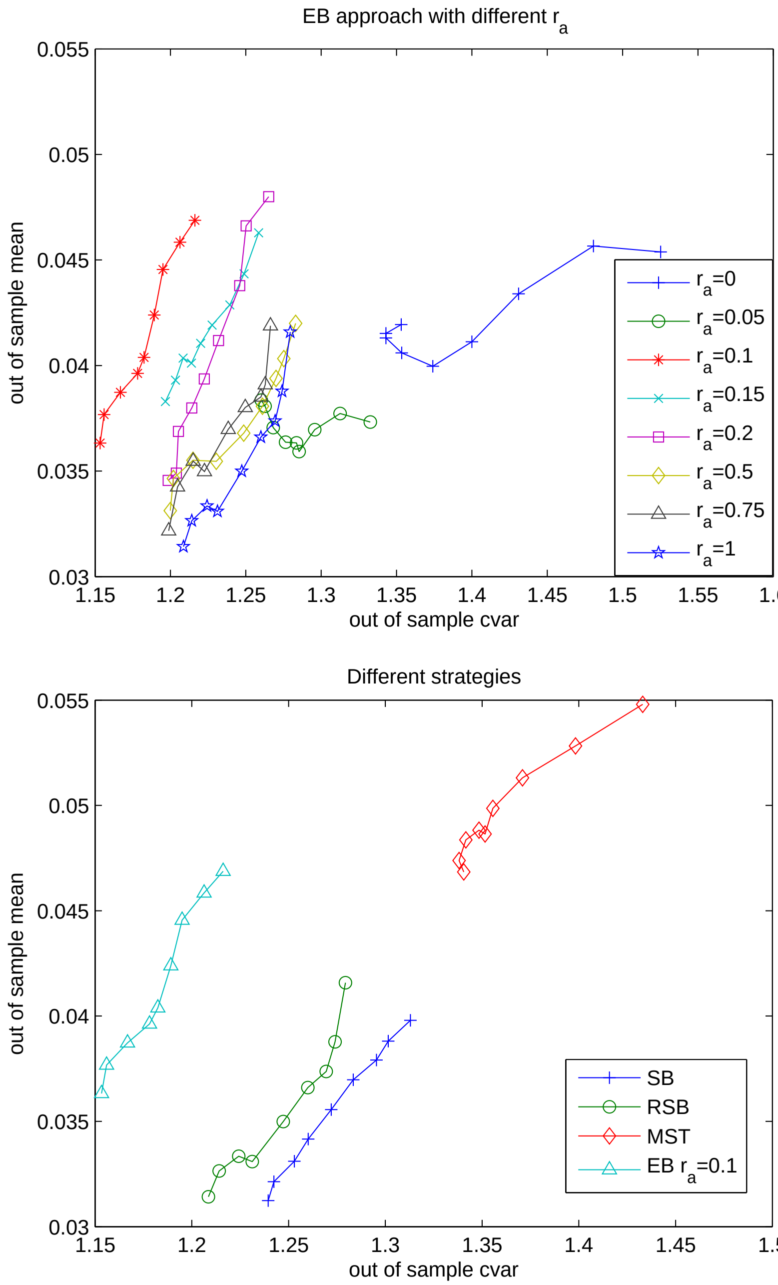

Figure 3: Out-of-sample efficiency frontier of different trading strategies, $\alpha = 0.9$

univariate marginals and in this example, its efficient frontier is much worse than those of the EB strategy with other values of r_a. This is to be expected since when $r_a = 0$, we use no dependency information from the financial market in the optimization model which is too weak. As r_a increases, the performance of the EB strategy improves, and the best efficient frontier is achieved when $r_a = 0.1$. The performance then gradually deteriorates as r_a continues to increase to 1. This indicates that using only partial dependency information can enhance the performance of the trading strategy in the out-of-sample data. The bottom graph shows the efficient frontiers of the four different strategies. The EB strategy is selected with the optimum value $r_a = 0.1$. We can see that EB and MST strategies are the best performing. While MST achieves higher out-of-sample mean, the CVaR values are higher too. Finally, the performance for the SB and RSB approaches are very close to each other and actually, the RSB trading strategy performs better than the SB one in this case. This serves to indicates that the rounding approach is reasonable. We repeat the experiment with $\alpha = 0.95$ and the results are displayed in Figure 4. We observe similar behavior in this case. The EB strategy with the best value is $r_a = 0.1$ in this case. The performance of EB and MST are better than the sample based approaches. The efficient frontiers of SB and RSB are close to each other, except that in this case, the performance of RSB is slightly worse than that of SB.

A well-known phenomenon in financial data is that the estimation of the out-of-sample mean is inaccurate (see Merton [27]). This is also true in our numerical experiments. The out-of-sample means are from 0.03 to 0.055, while the target returns are between 0.04 and 0.08. We conduct an experiment in which we directly use the out-of-sample mean data in the optimization formulation. While clearly impractical, this serves to check the effect of the inaccuracies in the estimation of the mean return on the comparative performance of the different strategies.

Figure 5 shows that in this case the EB strategy is the best performing strategy while the MST strategy does not perform as well. This seems to imply that the cover structure from the EB strategy is better that the spanning tree cover from the MST strategy. From these experiments, we conclude that the parameterized EB strategy achieves the best performance in this dataset. Note that out approach is completely data-driven from identifying the cover to computing the optimal portfolio. We conclude this section by showing an example of resulting covers from MST and EB strategy with $r_a = 0.1$. In this example, almost all of 48 two-element subsets of MST cover appears in the subsets of EB cover. This implies that there is greater dependence information used in the

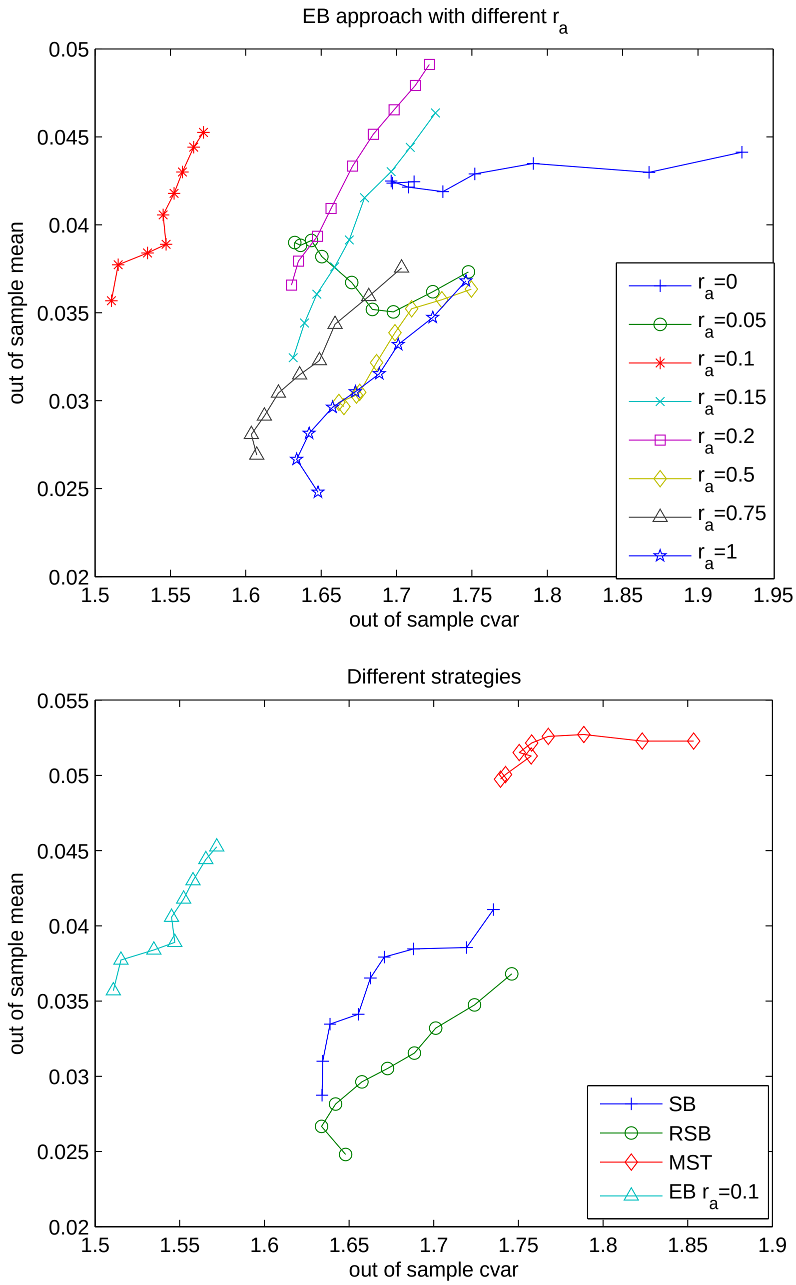

Figure 4: Out-of-sample efficient frontiers of different strategies, $\alpha = 0.95$

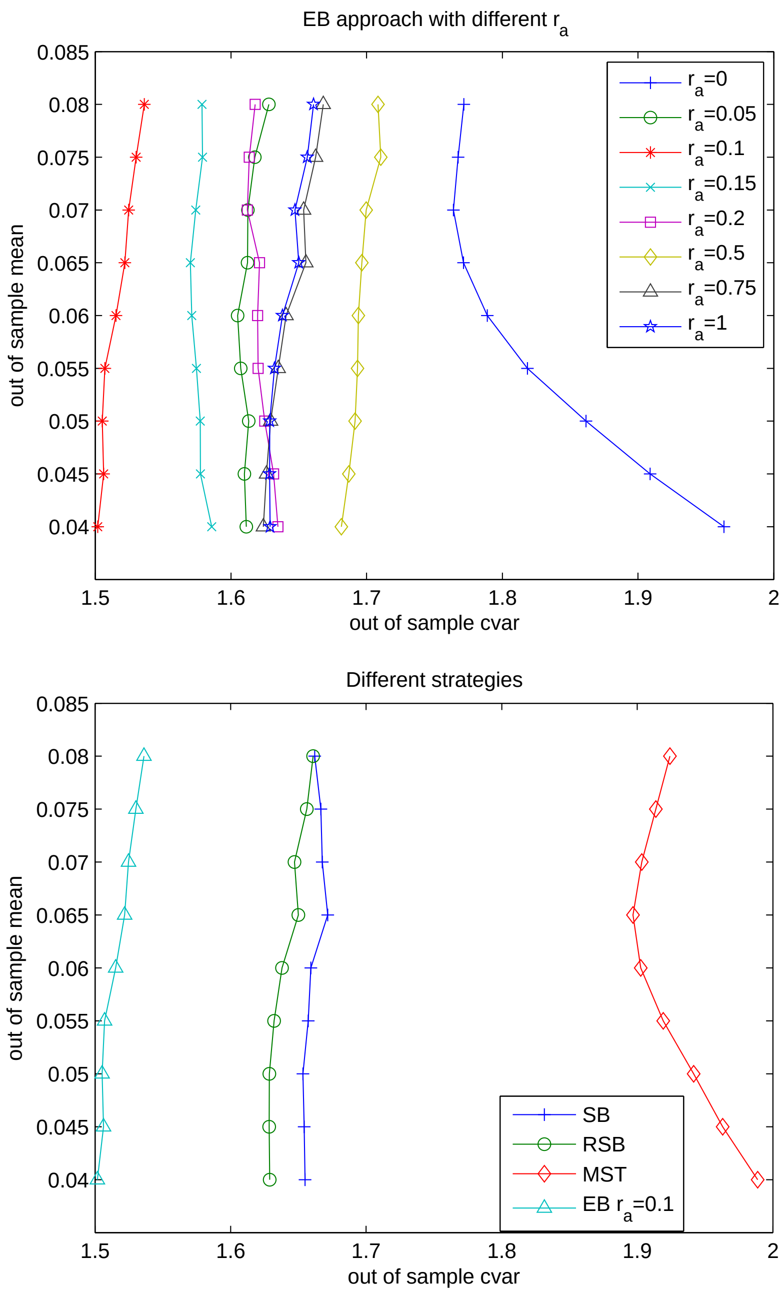

Figure 5: Out-of-sample efficient frontiers of different strategies, $\alpha = 0.95$, out-of-sample mean fixed

EB strategy, which adds value in our computations. The EB cover in this example has 37 subsets, with the largest subsets consisting of 13 elements. The top diagram in Figure 6 shows the first four subsets of this particular cover while the bottom diagram shows the 4th subset and five additional ones, from the 5th to the 9th, which is already much more complicated than the tree structure of the MST cover.

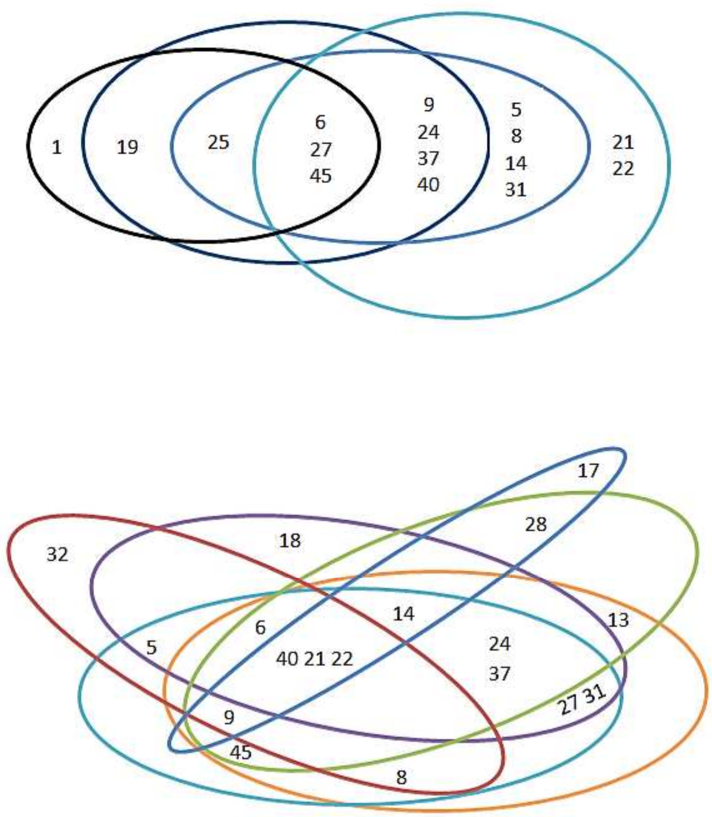

Figure 6: Partial Venn diagrams of an EB cover example with $r_a = 0.1$

5 Conclusion

In this paper, using the graph theoretic - running intersection property, we developed a linear program to compute Fréchet bounds on random portfolio risks. The formulation was shown to be efficiently solvable for the discrete distribution case. We proposed robust bounds on CVaR and VaR of the joint portfolio with overlapping multivariate marginal distribution information. Based on the tight and efficiently solvable bounds, we proposed a novel data-driven robust portfolio optimization model. The model identifies the overlapping structure of joint risks by computing the changes in correlation over time. We used historical data based on the identified overlapping multivariate marginal structure to optimize the portfolio allocation. The results suggest that by using only properly chosen partial distributional information, the out of sample performance can be enhanced.

References

[1] J. G. Shanthikumar A. E. B. Lim and G-Y. Vahn. Conditional value-at-risk in portfolio optimization: Coherent but fragile. *Operations Research Letters*, 39:163–171, 2011.

[2] D. Bertsimas, X. V. Doan, K. Natarajan, and C.-P. Teo. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3):580–602, 2010.

[3] K. Natarajan C. P. Teo. Bertsimas, D. Persistence in discrete optimization under data uncertainty. *Mathematical Programming, Series B, Issue on Opimization under Uncertainty*, 108(2-3):251–274, 2006.

[4] X. V. Doan and K. Natarajan. On the complexity of non-overlapping multivariate marginal bounds for probabilistic combinatorial optimization problems. *Operations Research*, 60(1):138–149, 2012.

[5] M. Dunham and K. Murphy. PMTK3: Probabilistic modeling toolkit for Matlab/Octave, version 3. `https://code.google.com/p/pmtk3/`, October 2012.

[6] P. Embrechts, A. Höing, and A. Juri. Using copulae to bound the value-at-risk for functions of dependent risks. *Finance and Stochastics*, 7(2):145–167, 2003.

[7] P. Embrechts and G. Puccetti. Bounds for functions of dependent risks. *Finance and Stochastics*, 10(3):341–352, 2006.

[8] P. Embrechts and G. Puccetti. Bounds for the sum of dependent risks having overlapping marginals. *Journal of Multivariate Analysis*, 101(1):177–190, 2010.

[9] M. Fréchet. *Les probabilités associées à un système d'événements compatibles et dépendants. I. Evenements en nombre fini fixe.* Actualités Sci. Ind., n. 859. Hermann et Cie, Paris, 1940.

[10] M. Fréchet. Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon, Section A, Series 3*, 14:53–77, 1951.

[11] K. French and E. Fama. This is a data library @online, 2013.

[12] L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Operations Research*, 51(4):543–556, 2003.

[13] C. Huang and A. Darwiche. Inference in belief networks: A procedural guide. *International Journal of Approximate Reasoning*, 15(3):225–263, 1996.

[14] CVX Research Inc. CVX: Matlab software for disciplined convex programming, version 2.0 beta. `http://cvxr.com/cvx`, September 2012.

[15] R. Jiroušek. Solution of the marginal problem and decomposable distribution. *Kybernetika*, 27(5):403–412, 1991.

[16] P. Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk.* McGraw-Hill, 2001.

[17] H. G. Kellerer. Verteilungsfunktionen mit gegebenen marginalverteilungen. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 3:247–270, 1964.

[18] H. G. Kellerer. Measure theoretic versions of linear programming. *Mathematische Zeitschrift*, 198:367–400, 1988.

[19] H. G. Kellerer. Indecomposable marginal problems. In G. D. Aglio, S. I. Kotz, and G. Salinetti, editors, *Advances in Probability Distributions with Given Marginals*, pages 139–149. Mathematics and Its Applications Volume 67, 1991.

[20] V. Kreinovich and S. Ferson. Computing best-possible bounds for the distribution of a sum of several variables is NP-hard. *International Journal of Approximate Reasoning*, 41(3):331–342, 2006.

[21] R. Uppal L. Garlappi and T. Wang. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, 20(7):41–81, 2007.

[22] J. B. Lasserre. Convergent sdp-relaxations in polynomial optimization with sparsity. *SIAM Journal of Optimization*, 17:833–843, 2006.

[23] S. L. Lauritzen, T. P. Speed, and K. Vijayan. Decomposable graphs and hypergraphs. *Journal of the Australian Mathematical Society*, 36:12–29, 1984.

[24] G. D. Makarov. Estimates for the distribution function of the sum of two random variables with given marginal distributions. *Theory of Probability and Its Applications*, 26:803–806, 1982.

[25] H. M. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.

[26] I. Meilijson and A. Nadas. Convex majorization with an application to the length of critical path. *Journal of Applied Probability*, 16:671–677, 1979.

[27] R. C. Merton. On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics*, 8:323–361, 1980.

[28] A. Natanzon, R. Shamir, and R. Sharan. A polynomial approximation algorithm for the minimum fill-in problem. *SIAM Journal on Computing*, 30(4):1067–1079, 2000.

[29] K. Natarajan, M. Sim, and J. Uichanco. Tractable robust expected utility and risk models for portfolio optimization. *Operations Research*, 20(4):695–731, 2009.

[30] K. Natarajan, M. Song, and C. P. Teo. Peristency model and its applications in choice modeling. *Management Science*, 55(3):453–469, 2009.

[31] G. Puccetti and L. Rüschendorf. Bounds for joint portfolios of dependent risks. *Statistics and Risk Modelling*, 29:107–132, 2012.

[32] G. Pucetti and L. Rüschendorf. Sharp bounds for sums of dependent risks. *Journal of Applied Probabililty*, 50(1):42–53, 2013.

[33] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2000.

[34] R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26:1443–1471, 2002.

[35] D. J. Rose, R. E. Tarjan, and G. S. Lueker. Algorithmic aspects of vertex elimination on graphs. *SIAM Journal on Computing*, 5(2):266–283, 2004.

[36] L. Rüschendorf. Random variables with maximum sums. *Advances in Applied Probability*, 14(3):695–731, 1982.

[37] L. Rüschendorf. Bounds for distributions with multivariate marginals. In K. Mosler and M. Scarsini, editors, *Stochastic Orders and Decision under Risk*, pages 285–310. IMS Lecture Notes - Monograph Series 19, 1991.

[38] L. Rüschendorf. Fréchet-bounds and their application. In G. D. Aglio, S. I. Kotz, and G. Salinetti, editors, *Advances in Probability Distributions with Given Marginals*, pages 151–187. Mathematics and Its Applications Volume 67, 1991.

[39] L. Rüschendorf. Stochastic ordering of risks, influence of dependence and a.s. constructions. In N. Balakrishnan, I. G. Bairamov, and O. L. Gebizlioglu, editors, *Advances on Models, Characterizations and Applications*, pages 19–56. CRC Press, 2005.

[40] N. N. Vorob'ev. Consistent families of measures and their extensions. *Theory of Probability and Its Applications*, 7:147–163, 1962.

[41] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1:1–305, 2008.

[42] B. Wang and R. Wang. The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, 102(10):1344–1360, 2011.

[43] H. Wang and M. Song. Ckmeans.1d.dp: optimal k-means clustering in one dimension by dynamic programming. *The R Journal*, 3(2):29–33, 2011.

[44] M. Yannakakis. Computing the minimum fill-in is NP-complete. *SIAM Journal on Algebraic and Discrete Methods*, 2(1):77–79, 1981.

[45] S. Zhu and M. Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, 57(5):1155–1168, 2009.

Appendix

Proof of Lemma 2

$\Rightarrow$. Since the cover is connected, for all $r \in \mathcal{R}\backslash\{1\}$, there exists a sequence $s_1 = r, s_2, \ldots, s_m = r-1$ that links r to $r-1$. If $s_2 < r$, we have: $K_r = J_r \cap \left(\bigcup_{t=1}^{r-1} J_t\right) \supseteq J_r \cap J_{s_2} \neq \emptyset$. If $s_2 > r$, there exist three consecutive indices in the sequence such that $s_{j-1} < s_j$ and $s_j > s_{j+1}$. We have: $J_{s_{j-1}} \cap J_{s_j} \neq \emptyset$ and $J_{s_{j+1}} \cap J_{s_j} \neq \emptyset$. Since $\mathcal{E}$ satisfies the RIP, we have $J_{\sigma_{s_j}} \supseteq K_{s_j} = J_{s_j} \cap \left(\bigcup_{t=1}^{s_j-1} J_t\right) \supseteq J_{s_j} \cap \left(J_{s_{j-1}} \cup J_{s_{j+1}}\right)$, thus we can replace s_j with σ_{s_j} in the sequence, with $\sigma_{s_j} < s_j$. Continuing on doing the process, we can find a sequence with $s_2 < r$.

$\Leftarrow$. We shall prove by induction on R. When $R = 2$, if $K_2 = J_1 \cap J_2 \neq \emptyset$, the cover is obviously connected. Suppose the statement is true for $R = k$, let us consider $R = k+1$. Since $K_r \neq \emptyset$ for all $r = 2, \ldots, k$, the subsets $J_1, \ldots, J_k$ are connected. Since $K_{k+1} \neq 0$, $J_{k+1} \cap J_{\sigma_{k+1}} \neq \emptyset$. Thus for all $r = 2, \ldots, k$, there exists a sequence linking r with $k+1$ via σ_{k+1}.

Numerical Example

Note that if there exists $i = 1, \ldots, N-2$ such that $d_i \leq 0$, the inequality $H(\boldsymbol{d}; x) \leq 0$ always holds since F_y and F_z are cumulative distribution functions. Thus we can restrict the feasible region to the positive orthant, $d_i > 0$ for all $i = 1, \ldots, N-2$. Similarly, we only need to consider solutions that satisfies $d_{N-1} = x - \sum_{i=1}^{N-2} d_i > 0$. Note that F_y and F_z are both continuously differentiable in

$(0, +\infty)$. Consider the first-order necessary optimality conditions, $\nabla H(\boldsymbol{d}; x) = \mathbf{0}$:

$$F'_y(d_1) = F'_y\left(x - \sum_{i=1}^{N-2} d_i\right) = F'_z(d_i), \qquad i = 2, \dots, N-2. \tag{30}$$

In order to solve this system of equations, we need the derivative F'_y and F'_z:

$$F'_y(d) = \begin{cases} 2d, & 0 < d < 1/2, \\ 1, & 1/2 \le d < 1, \\ -2d+3, & 1 \le d < 3/2, \\ 0, & d \ge 3/2, \end{cases} \quad \text{and} \quad F'_z(d) = \begin{cases} 4d, & 0 < d < 1/2, \\ -4d+4, & 1/2 \le d < 1, \\ 0, & d \ge 1. \end{cases}$$

We then need to consider (30) three distinct cases with $F'_y(d_1) = 0$, $0 < F'_y(d_1) < 1$, and $F'_y(d_1) = 1$.

1. $F'_y(d_1) = 0$: Since $d_1 > 0$, we have $d_1 \ge 3/2$. Similarly, we have $d_i \ge 1$ for all $i = 2, \dots, N-2$ and finally, $d_1 = x - \sum_{i=1}^{N-2} d_i > 3/2$. This case happens only when $x = \sum_{i=1}^{N-1} d_1 \ge (N-3) + 2 \times \dfrac{3}{2} = N$, which is the trivial case with $\mathrm{RBS}(x) = 1$ since $\sum_{i \in \mathcal{N}} \tilde{c}_i < N$ almost surely.

2. $0 < F'_y(d_1) < 1$: Let $z = F'_y(d_1)/4$, we have $z \in (0, 1/4)$ and $F'_y(2z) = F'_y(d_1)$. Given the formulation of $F'_y(d)$, we can easily show that $d_1 \in \{2z, 3/2 - 2z\}$. Similarly, we have: $d_{N-1} \in \{2z, 3/2 - 2z\}$. Now consider d_i, $i = 2, \dots, N-2$, we have $F'_z(z) = F'_z(1-z) = F'_z(d_i)$. Thus $d_i \in \{z, 1-z\}$ for all $i = 2, \dots, N-2$. Let k be the number of decision variables among d_i, $i = 2, \dots, N-2$ that take the value of z. We have: k can take any value from 0 to $N-3$. Similarly, let l be the number of decision variables among $\{d_1, d_{N-1}\}$ that take the value of $2z$, $l = 0, 1, 2$. Using the constraint $\sum_{i=1}^{N-1} d_i = x$, we obtain the following equation on z:

$$(N - k - 3l/2) - (N + 1 - 2k - 4l)z = x.$$

   If this equation results in a solution $z \in (0, 1/4)$, we achieve a set of solutions $\boldsymbol{d}$ of the original problem which satisfy the first-order optimality condition (30). It means we would need to consider $3(N-2)$ possible value pairs of (k, l) and check the feasibility of z to find all potential candidates of the optimal solution for this case.

3. $F'_y(d_1) = 1$: In this case, we have $d_1 \in [1/2, 1]$ and so is d_{N-1}. For $i = 2, \dots, N-2$, we have $d_i \in \{1/4, 3/4\}$. Similarly to the previous case, we let k be the number of decision variables

among d_i, $i = 2, \dots, N-2$ that take the value of $1/4$, we then have:

$$(3N - 9 - 2k)/4 + d_1 + d_{N-1} = x.$$

In addition, the objective value in this case can be computed as

$$H(\boldsymbol{d}; x) = d_1 + d_{N-1} - 1/2 + kF_z(1/4) + (N - 3 - k)F_z(3/4) - (N - 2).$$

Thus, we just need to check whether $y = d_1 + d_{N-1} = x - (3N - 9 - 2k)/4$ belong to the interval $[1, 2]$. We need to perform this feasibility check for $N - 2$ different values of k.

Following the analysis of these cases, we can find the optimal solution among all potential candidates, which will help us compute the standard bound $\mathrm{RSB}(x)$.